

# **Sky's the limit: trait-based modelling of food webs and migratory bird movement**

Dominique Caron

Biology Department  
McGill University, Montreal  
May, 2024

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of  
Doctor of Philosophy

*À maman, dont l'amour, la gentillesse et la confiance en moi m'ont soutenu et inspiré à chaque étape de ma vie.*

*À papa, qui, je sais, aurait été fier de moi.*

# Table of Contents

Acknowledgements.....	I
Abstract.....	III
Résumé.....	V
Table of Figures.....	VII
Index of Tables.....	VIII
Contribution to original knowledge.....	IX
Contributions of Authors.....	XI
Introduction.....	1
The ecological niche.....	1
Relations between species traits and niche.....	4
Biogeography.....	6
Trophic interactions and food webs.....	9
Seasonal bird movements and threats.....	12
Objectives.....	15
Literature cited.....	17
Chapter 1: Addressing the Eltonian shortfall with trait-based interaction models.....	31
Abstract.....	32
Introduction.....	33
Materials and methods.....	35
Results.....	43
Discussion.....	46
Aknowledgements.....	49
References.....	51
Table and figures.....	57
Bridging chapter 1 and chapter 2.....	62
Chapter 2: Trait-matching models predict pairwise interactions across regions, not food web properties .....	63
Abstract.....	65
Introduction.....	67
Methods.....	70
Results.....	78
Discussion.....	81
References.....	87

Data Accessibility Statement.....	96
Acknowledgements.....	96
Tables and figures.....	97
Bridging chapter 2 and chapter 3.....	103
Chapter 3: Consequences of seasonal geographic and niche movements for North American bird biogeography.....	104
Abstract.....	105
Introduction.....	106
Material and methods.....	110
Results.....	114
Discussion.....	116
References.....	122
Table and figures.....	130
Bridging chapter 3 and chapter 4.....	135
Chapter 4: Seasonal level of human impact and protection drive the decline of the North American avifauna.....	136
Summary.....	137
Introduction.....	138
Results.....	142
Discussion.....	145
Figures.....	149
STAR Methods.....	153
References.....	161
Discussion.....	173
Trait-based predictions of trophic interactions.....	176
The assembly of food webs.....	178
Seasonal bird biogeography.....	181
Appendices.....	186
A. Supplementary material for chapter 1.....	186
Table of Content:.....	186
Appendix S1: Amphibian body mass imputation.....	187
Appendix S2: Model runs and potential scale reduction factor.....	188
Appendix S3: Predictive model sensitivity analyses.....	190
Appendix S4: Predictor Importance.....	194
Appendix S5: Miscalibration results.....	195

Appendix S6: Bioregional models and food webs.....	196
Appendix S7: Additional performance metrics.....	199
<b>B. Supplementary material for chapter 2.....</b>	<b>203</b>
Table of Content:.....	203
Appendix S1: Description of the food webs.....	204
Appendix S2: Trait coverage and imputation.....	206
Appendix S3: Description of the structure of the predictive model.....	211
Appendix S4: Description of species roles and food web-level properties.....	215
Appendix S5: Boosted regression tree methods and results.....	219
Appendix S6: Convergence of the predictive models.....	224
Appendix S7: Predicting species interactions extended results.....	237
Appendix S8: Predicting species role extended results.....	242
Appendix S9: Predicting food web properties extended results.....	263
Appendix S10: Results with alternative Serengeti food web and without the Serengeti.....	264
Appendix S11 – References.....	272
<b>C. Supplementary material for chapter 3.....</b>	<b>276</b>
Table of Contents:.....	276
Appendix S1: Latitudinal and trait bias of species included in analyses.....	277
Appendix S2: Alternative metrics of movement in geographic and climate space movements.....	281
Appendix S3: Phylogenetic multilevel models.....	288
Appendix S4: Model convergence.....	296
<b>D. Supplementary material for chapter 4.....</b>	<b>303</b>
Appendix S1 – Multicollinearity sensitivity analysis.....	304

## Acknowledgements

This thesis was made possible by invaluable support of many individuals, for whom I am profoundly grateful.

I thank my supervisor Laura Pollock. You have been an incredible mentor throughout this endeavour. Your energy, kindness, patience, and intelligence made these five years (mostly) enjoyable and fulfilling. I am especially grateful for the time, space, and support you provided me during academic and personal challenges. You are truly a great human being, and I genuinely could not have asked for a better supervisor.

As Laura's first PhD student at McGill, I had the privilege of witnessing the lab grow into a group of dynamic, smart and fun researchers. I thank my lab mates: Julia McDowell, Isaac Eckert, Abbie Gail-Jones, Andrea Brown, Janaína Serrano, Nina Obiar, Samara Manzin, Noah Wightman, Maximilianne Jousse, Samantha Strauss, Katherine Hébert, and Wenyuan Zhang. You provided healthy distractions from work and alleviated the occasional solitude of the PhD experience.

I thank my numerous collaborators and committee members, especially: Brian Leung, Dominique Gravel, Timothée Poisot, Tanya Strydom, Michael Catchen, Wilfried Thuiller, Luigi Maiorano, Ulrich Brose, Miguel Lurgi, Gabriel Bergeron, Louis Moisan, Azénor Bideault, Guillaume Blanchet, Pierre Legagneux, Morgan Tingley, Dirk Karger, Alex Fuster, Philippe Archambault, and David Beauchesne. Each of you has contributed to varying extents to the development of the ideas presented in this thesis and to my growth as a researcher.

This work would not have been possible without funding and support from NSERC, the Quebec Centre for Biodiversity Science (QCBS), McGill, and the Canadian Institute for Ecology & Evolution (CIEE).

Lastly, I am grateful for the opportunity to pursue my PhD in my hometown, which allowed me to have continued support from my family and lifelong friends. À ma famille (papa, maman, Guillaume, Alexandre, Fanny, Léa et Maël), à mes amis du secondaire (Édouard, Antoine, David, François, Pierre-Alexis et Victor), à mes amis d'escalade (Bisson, Alexis, Paul, Sarah, Simon, Jean-Simon, Ned, Sara, Sam, Hoai-Nam, Marilee, Marianne, Adrien et Lucie), à Cléo, et à ma copine Laurence: Merci! Vous avez été au cœur des meilleurs moments de ces dernières années. Vous êtes la source de mon bonheur. Vous êtes ce qu'il y a de plus important. Je vous aime.

## **Abstract**

The ecological niche encompasses how species interact with their abiotic and biotic environment. In ecology, a fundamental assumption is that species traits correlate with their ecological function and niche, providing a crucial framework for predicting community structure. However, the success of trait-based predictions relies on general relationships between traits and Grinnellian (environment) or Eltonian (species interactions) niches across clades, spatial scales, and time. My thesis takes a macroecological approach to test generalities into the relationships between traits and different niche axes. Chapter 1 and 2 ask how traits explain predator-prey interactions among terrestrial vertebrates across large-scale food webs. My findings indicate a remarkable generality in trait-trophic interaction relationships within taxa and diverse food webs. This suggests promising prospects for using traits to predict predator-prey interactions in under-sampled ecosystems or future communities without current analogs. Chapter 3, I focus on the relationship between annual movement in geographic space and within climatic niche spaces among North American birds. I ask whether species traits explain this relationship and its consequences for seasonal bird biogeography. My findings gives a more mechanistic understanding of the seasonal rearrangement of bird diversity and traits across North America. Finally, in Chapter 4, I explore how traits and seasonal levels of human impact and protection explain the decline among North American birds. I find complex and sometimes contrasting effects between the level of human impact and protection on species trends in different seasons. My findings suggest that the migration seasons is especially crucial in explaining the decline of migrating species. This support the recent calls for better protecting the entire migratory network of species and reducing human pressures in stopovers and staging areas. Overall, by unraveling trait-based predictions, this research enhances our understanding on the potential and limitations of predictive methods of

biodiversity responses, critical considerations in the context of rapidly changing climate and shifting seasonal dynamics.

## Résumé

La niche écologique décrit la manière dont les espèces interagissent avec leur environnement abiotique et biotique. En écologie, une hypothèse fondamentale est que les traits des espèces sont associés avec leur fonction écologique et leur niche, fournissant un cadre crucial pour prédire la structure des communautés. Cependant, le succès des prédictions basées sur les traits repose sur la généralité des relations entre les traits et les différents axes de la niche à travers les groupes taxonomiques, l'espace et le temps. Ma thèse adopte une approche macroécologique pour tester ces généralités dans les relations entre les traits et différents axes de niche. Les Chapitres 1 et 2 se penchent sur la relation entre les traits des vertébrés terrestres et leurs interactions prédateur-proie dans des réseaux trophiques à grande échelle. Mes résultats indiquent une remarquable généralité dans les relations entre traits et interactions trophiques au sein des taxons et des réseaux trophiques d'écosystèmes divers. Cela suggère des perspectives prometteuses pour l'utilisation des traits des espèces afin de prédire les interactions prédateur-proie dans des écosystèmes sous-échantillonnés ou des communautés futures sans analogues actuels. Dans le Chapitre 3, je me concentre sur l'influence des traits sur la relation entre le mouvement annuel dans l'espace géographique et l'espace climatique chez les oiseaux nord-américains. Je m'intéresse aussi aux conséquence de ces relations entre les traits et mouvements des espèces pour la biogéographie saisonnière des oiseaux. Mes résultats offrent meilleure compréhension de la réorganisation saisonnière de la diversité et des traits des oiseaux à travers l'Amérique du Nord. Enfin, dans le Chapitre 4, j'explore comment les traits et les niveaux saisonniers d'impact humain et de protection expliquent le déclin des oiseaux de l'Amérique de Nord. Mes analyses indiquent des effets complexes et parfois contrastés entre le niveau d'impact humain et de protection sur les tendances des espèces à différentes saisons. Mes résultats suggèrent que les saisons de migration sont particulièrement cruciales pour expliquer le déclin des espèces migratrices. Cela supporte les appels

récents à mieux protéger l'ensemble du réseau migratoire des espèces et à réduire les pressions anthropiques dans les haltes migratoires. En somme, en s'intéressant aux prédictions basées sur les traits des espèces, cette recherche améliore notre compréhension du potentiel et des limites de ces méthodes pour prédire les réponses de la biodiversité, des considérations critiques dans le contexte du changement climatique rapide et des dynamiques saisonnières changeantes.

# Table of Figures

## Chapter 1:

Figure 1-1: Effect of the number of interactions sampled on the predictive performance of the model.	56
Figure 1-2: Differences in predictive performance among groups.....	57
Figure 1-3: Effect of generality on model performance.....	58
Figure 1-4: Spatial variation in predictability of local food webs.....	59

## Chapter 2:

Figure 2-1: Trophic interactions model transferability analysis workflow.....	96
Figure 2-2: Transferability of predictive models across geographic, environmental and phylogenetic distances.....	97
Figure 2-3: Predicting species interactions.....	98
Figure 2-4: Predicting species functional role.....	99
Figure 2-5: Prediction error of global food web properties.....	100

## Chapter 3:

Figure 3-1: Trade-off between movements in geographic and climatic space.....	129
Figure 3-2: Relationship between migratory behaviour and species traits.....	130
Figure 3-3: Seasonal variation in the latitudinal diversity gradient.....	131
Figure 3-4: Seasonal variation in Bergmann's rule.....	132

## Chapter 4:

Figure 4-1: Seasonal changes in bird abundances between 1970 and 2017.....	147
Figure 4-2: Weekly variation in level of human footprint (top) and protection (bottom) for migratory water birds (left) and migratory land birds (right).....	148
Figure 4-3: Effects of seasonal level of human footprint and protection on species trends.....	149
Figure 4-4: Effects of species traits on species trends.....	150

# Index of Tables

## Chapter 1:

Table 1-1: Variables used to predict trophic interactions between all species pairs. *The variables are grouped as foraging traits, vulnerability traits and matching traits. BM stands for body mass.....* 57

## Chapter 2:

Table 2-1: Summary statistics of the food webs used in this study after excluding species for which no match was found in the GBIF Backbone Taxonomy or was not documented in the trait database. Trait coverage is the percentage of traits documented in Etard et al. (2020) before imputing missing traits.. 97

Table 2-2: Area under the receiver operating curve (AUC) and area under the precision-recall-gain curve (AUPRG) where each food web model is used to predict food web data..... 97

## Chapter 3:

Table 3-1: Effect of each traits on migration distance (A) and climatic distance (B). Univariate regression columns show the total effect of each traits (not controlling for other traits), whereas multivariate regression columns show the direct effect of each traits (controlling for other traits). The values are the mean coefficients and the 95% credible interval from the posterior predictive distribution of the models..... 130

# **Contribution to original knowledge**

This manuscript-based thesis represents distinct contributions to original knowledge. All chapters are either published, or in preparation for submission. Chapter 1 is published in *Ecology Letters*, chapter 2 is published in *Global Ecology and Biogeography*, chapter 3 will be submitted to *Diversity and Distributions*, and chapter 4 will be submitted to *Current Biology*. The following are, to the best of my knowledge, the most novel aspects of this thesis:

## **Chapter 1:**

- I perform the first trait-based predictions of trophic interactions between terrestrial vertebrates, showing generalities in trait constraints across many taxa.
- This is the example of trophic interaction prediction at the largest scale in terms of number of species and space, showing the potential trait-based models to make first approximations of regional metawebs with limited data.

## **Chapter 2:**

- I demonstrate for the first time the transferability of predictive model of trophic interactions, showing the generalities in trait-relationships between terrestrial predators and prey across ecosystems.
- I show for the first time how predictive model of trophic interactions do not predict many food web properties, demonstrating the limit of these models to make network inferences.

## **Chapter 3:**

- I use for the first time bird distributions over their full annual cycle to relate bird traits, migration, and degree of ‘niche-tracking’.
- I explore for the first time the consequences of the relationship between species trait and migratory behaviour on ecogeographical rules.

#### **Chapter 4:**

- I show for the first time the relationship between human impact during bird migration and species trends across the entire assemblage of North American Birds.
- I explore for the first time the variations in level of protection and human impact over North American birds’ full annual cycle.

## Contributions of Authors

I am the sole first author on all chapters of this thesis. With the guidance of my supervisor, I developed the ideas of chapters, collated the data, conducted the analyses, produced the figures, wrote the first draft, and produced the final version of each manuscript.

*Chapter 1:* **Caron, D.**, Maiorano, L., Thuiller, W., & Pollock, L. J. (2022). Addressing the Eltonian shortfall with trait-based interaction models. *Ecology Letters*, 25(4), 889-899.

DC and LJP conceived the study, LM and WT provided comments on the design and analyses. DC collated the data and performed the analyses. DC wrote the first draft of the manuscript, and all authors contributed substantially to revisions.

*Chapter 2:* **Caron, D.**, Brose, U., Lurgi, M., Blanchet, F. G., Gravel, D., & Pollock, L. J. (2024). Trait-matching models predict pairwise interactions across regions, not food web properties. *Global Ecology and Biogeography*, e13807.

DC and LJP conceived the study, DG, GFB, ML, and UB provided comments on the design and analyses. DC collated the data and performed the analyses. DC wrote the first draft of the manuscript, and all authors contributed substantially to revisions.

*Chapter 3:* **Caron, D.**, Jousse, M., Tingley, M. W., Karger, D. N., Gravel, D., & Pollock L. J. (in prep). Consequences of seasonal geographic and niche movements for North American bird biogeography.

DC and LJP conceived the study. DC, DNK, and MJ collated the data. MWT, DNK, and DG provided comments on the design and analyses. DC performed the analyses. DC wrote the first draft of the manuscript, and all authors contributed substantially to revisions.

*Chapter 4: Caron, D., Legagneux, P., Beauchesne, D., Archambault, P., Pollock L. J. (in prep).*

Seasonal level of human impact and protection drive the decline of the North American avifauna.

DC and PL conceived the study. DB, PA, and LJP provided comments on the design and analyses. DC collated the data and performed the analyses. DC wrote the first draft of the manuscript, and all authors contributed substantially to revisions.

# Introduction

The goal of ecology, like other sciences, is to find generalities (MacArthur 1972). By seeking regularities, we can gain insights into the mechanisms shaping ecological communities and develop models to predict their behaviour under new conditions (Houlahan *et al.* 2017; McGill 2019). The overarching objective of my thesis is to find generalities on how species traits influence predator-prey interactions, migratory behaviour, and seasonal biogeography of terrestrial vertebrates. Specifically, I use a macroecological approach to ask: How general are trait relationships among predators and prey across all European terrestrial vertebrates (Chapter 1)? How general are predator-prey trait relationships among terrestrial vertebrates across ecosystems (Chapter 2)? How traits influence migratory behaviour of North American birds and what are the consequences for seasonal biogeography (Chapter 3)? And how traits and seasonal stressors drive North American bird trends (Chapter 4)? In this introduction, I will present five interrelated ecological concepts fundamental to these questions: the ecological niche, species traits, biogeography, trophic interactions, and seasonality. I will conclude the introduction by describing how these concepts tie into the overarching objectives and hypotheses of the entire thesis and each chapter.

## The ecological niche

Every species has requirements to persist in an environment. When these requirements are not met, and without sufficient immigration, populations are destined to local extinctions. This concept of environmental requirements is at the core of the ecological niche, defined as the biotic and abiotic factors for which a species has a positive growth rate (Chase & Leibold 2003; Holt 2009). Thus, the ecological niche encompass species response and impact within its environment, serving as a central concept for addressing some of ecology's most fundamental questions: Why are species found where

they are (Wiens 2011)? How do so many species coexist (Chesson 2000)? Why are some species rare and others common (Matthews & Whittaker 2015)? How do species interact with one another (Williams & Purves 2011)? What roles do species play in ecosystems (Cirtwill *et al.* 2018)? And, crucially, how will species and communities respond to rapid environmental changes (Tingley *et al.* 2009)?

From the outset, it was clear the ecological niche was a complex concept emerging from two diverging perspectives. Joseph Grinnell (1917) emphasized on the species relationships with abiotic factors in his seminal work on the distribution of the California Thrasher. He proposed that “[..] no two species regularly established in a single fauna have precisely the same niche relationships”, formulating the competitive exclusion principle later predicted by Vito Volterra’s (1928) mathematical models and demonstrated in lab experiments by Georgy Gause (1934). In contrast, Charles Elton (1927) focused on the position of species within its biotic environment, defining the niche as the “status of the animal in the community”. These two perspectives set the stage for G. Evelyn Hutchinson to develop his hypervolume model of the niche (Hutchinson 1957). For Hutchinson, the niche is the location where a species can establish and persist in a n-dimensional space where axes are environmental variables. His framework successfully integrate the Grinnellian dimensions of the niche (environmental requirements) in an intuitive quantitative framework, shaping the theoretical work in coexistence (Chesson 2000; Letten *et al.* 2017) and biogeography (Godsoe *et al.* 2017; Guisan & Zimmermann 2000) for decades.

The dynamic nature of Eltonian’s dimensions of the niche (relationships to other species) posed challenges in integrating them into Hutchinson’s hypervolume framework (McInerny & Etienne 2012). Recent research has focused on integrating the Grinnellian and Eltonian niches to better understand the distributions of species (Soberón 2007) and interactions (Gravel *et al.* 2019). This integration raised

new debates such as what is the role of species interactions on species distributions across scales? Traditionally, it was believed that species interactions primarily influence distributions at local scales, while climate shape regional and geographic scale distributions (Keddy 1992; Soberón 2007; Whittaker *et al.* 2001). However, there are now evidence that the effects of local species interactions can scale-up to influence distributions at all scales (Araújo & Rozenfeld 2014; Belmaker *et al.* 2015). Another debate that raised from attempts to integrate the Grinnellian and Eltonian niches is whether variation in interactions over environmental gradients is mainly driven by turnover of species or of interactions (Poisot *et al.* 2012; Tylianakis & Morris 2017). We know that species abundances and interactions can vary with the environment, but it is unclear what is the relative importance of these two mechanisms on variation of ecological networks (Gravel *et al.* 2019). These two debates highlight how a better integration of the different niche axes would help tackle some fundamental ecological questions.

Ecological communities are not only shaped by the deterministic processes of niche theory, but also by stochastic elements. In fact, neutral theory, first proposed by Hubbell (2001), posits that biodiversity is structured in space and time not by ecological differences among species (growth, death, dispersion rates) but by stochastic elements (ecological drift) alone. Despite the unrealistic assumption of perfect equivalence among species, neutral theory successfully predicts many community properties (Bell & Schluter 2000; Canard *et al.* 2012). Modern ecological theories now recognized both niche and neutral processes as central to shaping ecological communities (Vellend 2010), and playing crucial roles in explaining ecological succession, dynamics and interactions (Gravel *et al.* 2006; Morales-Castilla *et al.* 2015; Poisot *et al.* 2015).

The ecological niche is a central theme in most of the chapters of my thesis. In Chapters 1 and 2, I compare the Eltonian niches of terrestrial vertebrates across different ecosystems, relating species traits

to explain predator-prey relationships. In Chapter 3, I investigate the Grinnellian niche of North American birds in the context of seasonal environments, examining how birds move within their climatic niche space throughout their annual cycle and how this movement relates to migration.

## **Relations between species traits and niche**

A fundamental challenge with the concept of the niche is that it they cannot be easily quantified. Instead, we have to rely on other species characteristics, such as functional traits, that are related to species niches (McGill *et al.* 2006). Species functional traits are morphological, physiological or phenological characteristics of a species that influences its performance (growth, survival, reproduction) in an environment (Violle *et al.* 2007). These traits evolved in response to environmental conditions and interactions with other species (Reich *et al.* 2003). Although the concept of functional trait have originally been developed for plant ecology (Funk *et al.* 2017), there is ample evidence of functional traits relevant to animal responses to (Vandewalle *et al.* 2010) and their impact on (e.g., Lundgren *et al.* 2024) their environment. Therefore, functional traits provide insights into the role of niche differentiation and environmental filtering in species distributions and community assembly (Zakharova *et al.* 2019). For example, Lamanna *et al.*, (2014) showed how quantifying the volume, packing, and overlap of functional trait space across scales could test different hypotheses for the latitudinal gradient in species richness. A clear benefit of traits is that, unlike species niches, they can be directly measured, providing an accessible quantitative framework for prediction (Kearney *et al.* 2010).

Given that species with similar traits are expected to share similar responses to the environment, traits hold promise for improving predictions of species distributions and anticipating responses to

environmental changes. For example, using a hierarchical species distribution model, Pollock et al. (2012) showed how Eucalypt species' leaf area, plant height, and seed mass influenced species response to environmental gradients. Not only do these trait-based predictions provide a more mechanistic understanding of how traits influence species distributions, they have been shown to help model transfer across species and space (Vesk *et al.* 2021). Functional traits can also predict species vulnerability to global warming (Pacifici *et al.* 2017) and precipitation regimes (Griffin-Nolan *et al.* 2018). Thus, trait-based predictions serve as valuable tools for filling gaps in our understanding of species and ecosystems' distributions and responses to abiotic factors, particularly for those about which we have limited knowledge.

Species sharing traits should also share similar interactions. For example, in a study looking at all predator-prey interactions for European vertebrates, O'Connor et al. (2020) found that species traits strongly explained species trophic groups (species sharing similar resources and enemies). Therefore, traits also provide a framework to predict interactions (Bartomeus *et al.* 2016; Gravel *et al.* 2013). For trophic interactions, this means that the traits of predators should determine the type of species it can feed on. Similarly, the traits of the prey should determine the type of predator a species has. These assumptions form the foundation of 'trait-matching' models that aim to predict functionally feasible interactions within a species pool (Morales-Castilla *et al.* 2015). These trait-matching models have been shown to predict accurately potential species interactions for a number of ecosystems, taxa, and interaction types (Brousseau *et al.* 2018; Laigle *et al.* 2018; Pichler *et al.* 2020; Pomeranz *et al.* 2019). Once again, these trait-based models offer a promising avenue for gaining a mechanistic understanding of how traits relate to niches and for filling knowledge gaps regarding species interactions (Hortal *et al.* 2015).

Each chapter of my thesis examines the relationship between species traits and their ecology. In the first two chapters, I use trait-based models to predict trophic interactions, focusing on how species traits relate to their Eltonian niches. In Chapter 3, I use traits to predict how North American birds move within climatic and geographic spaces, and exploring the consequences for the seasonal redistribution of traits in space. Finally, in Chapter 4, I investigate how bird traits predict the decline among North American birds between 1970 and 2017.

## Biogeography

One of the subfield of ecology most profoundly influenced by niche theory is biogeography, the study of how biodiversity is distributed. The history of biogeography can be traced back to the work of Alexander von Humboldt. In his 1807 essay, Humboldt noted how plant communities tend to be more diverse as one moves closer to the equator (Humboldt *et al.* 2008). Mechanisms explaining this latitudinal gradient of diversity would be proposed only centuries after Humboldt's work and remains heavily debated (Mittelbach *et al.* 2007; Pianka 1966). Another key figure in the foundation of biogeography is Alfred Russel Wallace. During his extensive travels in the mid-19th century, Wallace noticed that vertebrate families are distributed tend to form distinct geographical units. Based on this observation and on the existing knowledge of his time, he proposed that the Earth's land surface can be divided in zoogeographic regions based on species distributions (Holt *et al.* 2013; Wallace 2011). These two seminal works paved the way for the exploration of other generalities and peculiarities of how the diversity of life distributes in space and time, such as the relationship between diversity and area (Preston 1960) or the work on island biogeography (MacArthur & Wilson 2001). Understanding how these patterns emerge have been among the most fruitful avenues to enhance our understanding of how communities assemble, persist and function.

The development of niche theory and models transformed biogeography from a primarily descriptive science into one that offers testable predictions (Elith & Leathwick 2009; Pulliam 2000). For example, species distribution models (SDMs), aim to quantify species niches by relating observed species occurrences to environmental variables (Elith & Leathwick 2009). However, the true relationship between species niche and what SDMs describe is correlative, and therefore often unclear. While SDMs depend on observations of the realized niches of species (Booth *et al.* 1988), they frequently fail to account for other factors influencing species distributions, such as biotic interactions and dispersal limitations (Araújo & Guisan 2006; Godsoe *et al.* 2017). This deficiency has been one of the main critique of SDMs, particularly regarding their spatial and temporal extrapolation (Thuiller *et al.* 2013; Urban *et al.* 2016; Zurell *et al.* 2009). Efforts have been made to integrate dispersal limitations and biotic interactions directly or indirectly into models, although implementation remains challenging due to data constraints (Engler & Guisan 2009; Pollock *et al.* 2014; Shipley *et al.* 2022; Staniczenko *et al.* 2017). Despite these limitations, SDMs have proven indispensable in bridging knowledge gaps, explaining biogeographic patterns, and emerging as a vital conservation tool (Guisan *et al.* 2013; Pollock *et al.* 2020).

Some biogeographic patterns are so widespread that they have been designated as ecogeographic “rules”. One of these rules is the latitudinal diversity gradient, the same decrease in species diversity towards the poles that Humboldt noted (Hillebrand 2004). Various explanations have been proposed for this gradient, including the effects of higher productivity and climate stability in the tropics on speciation rates, niche partitioning, specialization, and biotic interactions (Mittelbach *et al.* 2007; Pontarp *et al.* 2019). Another ‘rule’ of longstanding interest is the relationship between body mass and temperature. Originally proposed by Bergmann (1848), the ‘rule’ is traditionally defined as ‘races of

warm blooded vertebrates from cooler climates tend to be larger than races of the same species from warmer climates' (Mayr 1956). While the exact reasons for this gradient's emergence are still debated, it is generally believed that larger-bodied animals possess better heat retention, enabling them to thrive in colder environments compared to smaller individuals (Blackburn *et al.* 1999). This principle has been extended to encompass the relationship between size and latitude for ectotherms and across closely related species, as originally proposed by Bergmann (Blackburn *et al.*, 1999).

A recent development in biogeography was to expand its focus from solely considering the distribution of organisms to trying to understand distribution of interactions and the ecological network they form (Windsor *et al.* 2023). Interaction networks can vary through two main processes: (1) changes in community composition, and (2) changes in the functional relationships between interacting species, like the body mass ratio between predators and prey (Poisot *et al.* 2015). These two processes are, in turn, influenced by factors such as the environment (Pellissier *et al.* 2018), the regional pool of interactions (Saravia *et al.* 2022), area (Galiana *et al.* 2018), or higher-order interactions (Poisot *et al.* 2015). Based on previous knowledge on species biogeography and on how species interaction network are structured, it is possible to formulate predictions regarding how network properties vary across space and scale (e.g., Baiser *et al.* 2019; Gravel *et al.* 2011) but the empirical investigation on these patterns remains rare (e.g., Baiser *et al.* 2012; Galiana *et al.* 2021; Gravel *et al.* 2019; O'Connor *et al.* 2020).

In this thesis, I investigate the biogeography of predator-prey interactions and bird diversity, traits, and trends across seasons. In Chapters 1 and 2, I examine how the trait relationships between predators and prey among terrestrial vertebrates vary across space. In Chapters 3 and 4, I use weekly bird

distributions to explore how bird movement within geographic and climatic spaces influences seasonal diversity, trait variations, and abundance changes across North America.

## Trophic interactions and food webs

In his definition of the niche, Elton focused on a species relationships with its resources and enemies (Elton 1927). These relationships, known as trophic interactions, involve a predator species feeding on a prey species. The collection of these trophic interactions among species form a food web, a fundamental concept in ecology to understand ecosystem structure and dynamics (Lindeman 1942; Thompson *et al.* 2012). The identity, strength and organization of trophic interactions influence how species distribute in space (Wisz *et al.* 2013), how populations respond to disturbances (Eklöf & Ebenman 2006), and how energy flow within and between ecosystems (Gounand *et al.* 2018). Given that food webs integrate biodiversity and ecosystem function, conserving their structure and the underlying trophic interactions is essential to protect the diversity and integrity of ecosystems (Harvey *et al.* 2017). For example, investigating the position of a species within a food web inform on its role within the community and help anticipate secondary extinctions (Cirtwill *et al.* 2018; McDonald-Madden *et al.* 2016).

Food webs have emerging properties with important implications on the communities, such as the level of connectivity (connectance), the structure of specialization among species, or the formation of modules (Delmas *et al.* 2019; Pimm *et al.* 1991). May famously showed that under random interaction networks, a more complex food web (measured by level of connectance, species richness, and mean interaction strength) should be less stable, challenging the commonly accepted assumption of the time that more complex systems are more stable (May 1972). This insight inspired decades of research

aimed at understanding how observed properties of empirical food webs, such as the organization of interactions, the distribution of interaction strengths or the correlation between interspecific interactions, influence the relationship between complexity and stability (Allesina & Tang 2015). Another example of how food web structure influences communities is demonstrated by network complementarity, where species partition resources and enemies, leading to increased ecosystem productivity (Poisot *et al.* 2013). Thus, it is not only the number and identity of trophic interactions that matters, but also how they are organized among species.

Yet, we still face major challenges when developing accurate descriptions of natural food webs and apply our theoretical understanding of food webs to conservation. One major obstacle is the lack of trophic interaction data across most locations and taxa (Poisot *et al.* 2021), which presents fundamental technical and practical challenges in food web ecology. In fact, the “lack of knowledge about interactions among species or among groups of species” remains one of the major biodiversity data shortfalls (Hortal *et al.* 2015). Sampling interactions is inherently difficult; observing an interaction requires simultaneously detecting individuals of two species while they are interacting. It is virtually impossible to sample all interactions even for a simple community and large sampling effort (Chacoff *et al.* 2012; Jordano 2016; Pringle & Hutchinson 2020). Additionally, there are spatial biases in available food web datasets, with a disproportionate focus on the United States and Europe (Cameron *et al.* 2019; Poisot *et al.* 2021). These biases collectively limit the scale of food web research and pose significant hurdles to the necessary shift in focus from species to interaction networks in conservation efforts (Harvey *et al.* 2017).

To address these gaps, models have been developed to predict trophic interactions based on food web theory (Strydom *et al.* 2021). Both neutral and niche processes determine which two species will

interact (Morales-Castilla *et al.* 2015). Neutral models of species interactions predict that feasible interactions are only dependent on species presences and abundances (Canard *et al.* 2012). The interactions that do not occur under this constraint are termed *neutrally forbidden interactions* (Morales-Castilla *et al.* 2015). Although neutral models can predict a significant amount of variation in interaction strengths (Canard *et al.* 2014), recent research has highlighted that co-occurrences do not always imply interaction (Blanchet *et al.* 2020; Thurman *et al.* 2019). In contrast to the neutral model, niche-based models predict that interactions between two species depend on their respective traits (Rossberg *et al.* 2010). Non-interactions driven by niche-based processes are termed *functionally forbidden interactions* (Morales-Castilla *et al.* 2015), and can be predicted by trait-based models of interactions (Bartomeus *et al.* 2016). Phylogenetic relationships also provide valuable information, serving as proxies for trait relationships, and because interactions and species' roles (i.e. species' positions in the food web) tend to be evolutionarily conserved (Gómez *et al.* 2010; Stouffer *et al.* 2012).

At the network level, factors beyond those acting at the interaction level (neutral and niche processes) also constrain food webs. Firstly, local interactions are drawn from the pool of interactions at the regional scale, thus local food webs are constrained by the regional web of interactions, often referred to as the metaweb. Recently, Saravia *et al.* (2022) showed that properties of food webs emerge from the metaweb, suggesting that the regional food web structure is primarily constrains local food web structure. Secondly, there are evidence that the environment influences the local realization of food webs. For example, primary productivity should determine the number of trophic levels and number of species at each trophic level an ecosystem can sustain (Thompson & Townsend 2005). Additionally, area have also been showed to influence the shape of the food webs (Galiana *et al.* 2018). Spatial (e.g., habitat diversity) and temporal (e.g., seasonality) factors are also believed to shape modules within

network (McMeans *et al.* 2015; Rooney *et al.* 2008). For example, species occurring in the same habitat type or season will form highly connected module relative to species not present at the same time or space. Finally, there are dynamic constraints on the local realization of food webs (Grilli *et al.* 2017). Theory predicts that for a food web to persist, it needs to lead to a feasible and stable equilibrium. In short, food webs are a product of neutral and niche constraints acting at the level of interactions, as well as neutral, environmental, and dynamic constraints at the network level.

Trophic interactions and the food webs they form are central themes of my thesis, particularly in Chapters 1 and 2. In these chapters, I investigate the generality of trait relationships between terrestrial vertebrate prey and predators. Additionally, in Chapter 2, I explore whether trait-based models of interactions predict food web properties, examining the importance of higher-level constraints (e.g., environmental and dynamic constraints) in structuring food webs. In Chapters 3 and 4, trophic interactions play a more indirect role. Here, I test how bird diet influences their movement within climatic and geographic spaces, as well as their changes in abundance.

## **Seasonal bird movements and threats**

From the beginning, Hutchinson highlighted the importance of seasonality (i.e., the annual cycle of environmental variables) in understanding species niches and coexistence (Hutchinson 1957, 1961). Despite the critical role of temporal variations, seasonality is often poorly integrated into most niche models (Ponti & Sannolo 2023). This is partly due to the inherent complexity of reconciling the niche concept when both the environment and species' responses to it fluctuate over time. Adding to this complexity, migrating species are moving both within climatic and geographic space, dynamically influencing their niches.

Migration is thought to have evolved in response to seasonality, allowing species to maintain fidelity to favorable breeding grounds while escaping harsh conditions (Winger *et al.* 2019). Unlike other adaptation to seasonality such as hibernation or freeze tolerance, migration allows species to maintain consistent thermal conditions throughout the year but comes with movement costs (Gómez *et al.* 2016; Nakazawa *et al.* 2004). Indeed, Somveille *et al.* (2015, 2019) showed that breeding and wintering destinations of migratory birds are optimized to avoid unfavorable climatic conditions, competition, and minimize travel distance. Therefore, evidence suggests there is a fundamental trade-off between the movement within geographic and climatic niche spaces (Gómez *et al.*, 2016; Nakazawa *et al.*, 2004), but it is poorly understood how the differences between species alter the balance of this trade-off.

Species' strategies for minimizing movement within their climatic niche space versus enduring variation in climatic conditions are influenced by geographic, morphological, life history, predatory, and feeding habits (Cohen & Jetz 2022; McKinnon *et al.* 2010; Zurell *et al.* 2018). For example, larger species should be more likely to minimize travel distance as they usually need more energy to migrate (Watanabe 2016) and can tolerate colder temperatures (Blackburn *et al.* 1999). Similarly, species with a flexible diet, like the Northern cardinal (*Cardinalis cardinalis*), which switches from feeding on insects and berries in summer to a mainly granivorous diet in winter (Gill 2007), are less likely to migrate to track resources. While these examples illustrate how certain traits explain why some birds favour movement within climatic space over movement in geographic space, we are still lacking evidence for the influence of other traits.

Understanding how migratory species interact within their niches is crucial as migratory species are facing significant declines worldwide (Bairlein 2016; Studds *et al.* 2017; Wilcove & Wikelski 2008). In

North America, Rosenberg et al. (2019) estimated a 29% decrease in bird abundance between 1970 and 2019. To mitigate this decline, we need a comprehensive understanding of bird ecology and stressors over their full annual cycle and protect their entire migratory network (Marra *et al.* 2015; Xu *et al.* 2020). Stressors impacting individuals in one season can carry-over to affect the survival and reproductive success at other seasons (Norris *et al.* 2004; Reudink *et al.* 2009). Despite the importance of a year-around understanding, most studies historically focused on a single season, usually the growing season (Marra *et al.* 2015).

As migratory birds change their ranges between seasons, so are the level of protection and the threats they experience. Runge et al. (2015) showed that globally, only 9% of 1451 migratory birds are adequately covered by protected areas across all stages of their annual cycle, compared with 45% for nonmigratory birds. Most Important Bird and Biodiversity Areas, which serve as the basis for bird site-based protection worldwide, are identified in the bird's breeding distributions. Therefore, it is not surprising that protection of migrating routes are often incompletely protected. Yet, the importance of adequate protection of the entire migratory network has been demonstrated and proposed as key in slowing down the decline of migratory species (Marra *et al.* 2015; Runge *et al.* 2015; Wilcove & Wikelski 2008; Xu *et al.* 2020; Zhang *et al.* 2023). Birds also experience different level of stress throughout their annual cycle. For example, La Sorte et al. (2022) recently showed that exposure to air pollution differ between seasons for world's nocturnally migrating birds, and that this exposure will also change differently among seasons in the future. Therefore, conservation outcomes depends on a deeper understanding of the level of protection and threats among bird species and how they are associated with the observed trends in abundances.

In my thesis, seasonality is a central concept in Chapters 3 and 4. In Chapter 3, I ask how traits influence the seasonal movements of birds in both geographic and climatic spaces. I also explore the seasonal variation in two biogeographic patterns among North American birds: the latitudinal diversity gradient and temperature-body mass relationship. In Chapter 4, I investigate the impact of human activities and the level of protection during different seasons on species trends. By examining these temporal dynamics, I aim to provide a deeper understanding of how seasonality influences niches of migratory species and highlight the importance of considering the full annual cycle of species in maintaining diversity.

## **Objectives**

The overarching objective of my thesis is to uncover generalities regarding how species traits explain ecological processes like predator-prey interactions and seasonal movements among terrestrial vertebrates. This research aims to better understand fundamental principles governing the biogeography of food webs and the seasonal rearrangement of species and traits. The central hypothesis guiding this work is that species traits play a critical role in shaping ecological niches, ultimately influencing the distribution, abundance, and persistence of species.

Each chapter of this thesis contributes to a deeper understanding of the relationships between species traits and ecological processes. In Chapter 1, my objective is to test how general are the trait relationships between predators and prey among European terrestrial vertebrates. Specifically, I explore how well traits predict trophic interactions across taxonomic groups and bioregions. In Chapter 2, I expand on this work by asking how general trait-interaction relationships are across four food webs of terrestrial vertebrates from contrasting environments. By doing so, I test the influence of the

environment on trait relationships between predators and prey. In this second Chapter, I also aim at testing whether trait-based models of interactions predict food web properties, asking whether trait constraints on interactions are the main drivers of variation of food web properties. In Chapter 3, my objective is to test the relationships between bird traits, and their seasonal movements in geographic and climatic space. By doing so, I explore the role of traits in explaining the seasonal biogeography of birds. Finally, in Chapter 4, I aim to test the importance of species traits and the seasonal levels of human impact on birds changes in abundances.

By integrating concepts of ecological niche, species traits, biogeography, trophic interactions, and seasonality, this thesis seeks to advance our knowledge of the mechanisms shaping ecological communities, provide tools to fill knowledge gaps on species interactions, and inform conservation efforts in a context of changing seasonal dynamics.

## Litterature cited

- Allesina, S. & Tang, S. (2015). The stability–complexity relationship at age 40: a random matrix perspective. *Population Ecology*, 57, 63–75.
- Araújo, M.B. & Guisan, A. (2006). Five (or so) challenges for species distribution modelling. *J Biogeography*, 33, 1677–1688.
- Araújo, M.B. & Rozenfeld, A. (2014). The geographic scaling of biotic interactions. *Ecography*, 37, 406–415.
- Bairlein, F. (2016). Migratory birds under threat. *Science*, 354, 547–548.
- Baiser, B., Gotelli, N.J., Buckley, H.L., Miller, T.E. & Ellison, A.M. (2012). Geographic variation in network structure of a nearctic aquatic food web: Network structure in an aquatic food web. *Global Ecology and Biogeography*, 21, 579–591.
- Baiser, B., Gravel, D., Cirtwill, A.R., Dunne, J.A., Fahimipour, A.K., Gilarranz, L.J., *et al.* (2019). Ecogeographical rules and the macroecology of food webs. *Global Ecology and Biogeography*, 28, 1204–1218.
- Bartomeus, I., Gravel, D., Tylianakis, J.M., Aizen, M.A., Dickie, I.A. & Bernard-Verdier, M. (2016). A common framework for identifying linkage rules across different types of interactions. *Functional Ecology*, 30, 1894–1903.
- Bell, G. & Schlüter, A.E.D. (2000). The Distribution of Abundance in Neutral Communities. *The American Naturalist*, 155, 606–617.
- Belmaker, J., Zarnetske, P., Tuanmu, M.-N., Zonneveld, S., Record, S., Strecker, A., *et al.* (2015). Empirical evidence for the scale dependence of biotic interactions. *Global Ecology and Biogeography*, 24, 750–761.

- Bergmann, C. (1848). *Über die Verhältnisse der Wärmeökonomie der Thiere zu ihrer Grösse*. Vandenhoeck und Ruprecht.
- Blackburn, T.M., Gaston, K.J. & Loder, N. (1999). Geographic gradients in body size: a clarification of Bergmann's rule. *Diversity and Distributions*, 5, 165–174.
- Blanchet, F.G., Cazelles, K. & Gravel, D. (2020). Co-occurrence is not evidence of ecological interactions. *Ecology Letters*, 23, 1050–1063.
- Booth, T.H., Nix, H.A., Hutchinson, M.F. & Jovanic, T. (1988). Niche analysis and tree species introduction. *Forest Ecology and Management*, 23, 47–59.
- Brousseau, P.-M., Gravel, D. & Handa, I.T. (2018). Trait matching and phylogeny as predictors of predator–prey interactions involving ground beetles. *Functional Ecology*, 32, 192–202.
- Cameron, E.K., Sundqvist, M.K., Keith, S.A., CaraDonna, P.J., Mousing, E.A., Nilsson, K.A., et al. (2019). Uneven global distribution of food web studies under climate change. *Ecosphere*, 10, e02645.
- Canard, E., Mouquet, N., Marescot, L., Gaston, K.J., Gravel, D. & Mouillot, D. (2012). Emergence of Structural Patterns in Neutral Trophic Networks. *PLOS ONE*, 7, e38295.
- Canard, E.F., Mouquet, N., Mouillot, D., Stanko, M., Miklisova, D. & Gravel, D. (2014). Empirical Evaluation of Neutral Interactions in Host-Parasite Networks. *The American Naturalist*, 183, 468–479.
- Chacoff, N.P., Vázquez, D.P., Lomáscolo, S.B., Stevani, E.L., Dorado, J. & Padrón, B. (2012). Evaluating sampling completeness in a desert plant–pollinator network. *Journal of Animal Ecology*, 81, 190–200.
- Chase, J.M. & Leibold, M.A. (2003). *Ecological Niches: Linking Classical and Contemporary Approaches*. University of Chicago Press.

- Chesson, P. (2000). Mechanisms of Maintenance of Species Diversity. *Annual Review of Ecology and Systematics*, 31, 343–366.
- Cirtwill, A.R., Dalla Riva, G.V., Gaiarsa, M.P., Bimler, M.D., Cagua, E.F., Coux, C., et al. (2018). A review of species role concepts in food webs. *Food Webs*, 16, e00093.
- Cohen, J. & Jetz, W. (2022). *Diverse strategies for tracking seasonal environmental niches at hemispheric scale* (preprint). *Ecology*.
- Delmas, E., Besson, M., Brice, M.-H., Burkle, L.A., Riva, G.V.D., Fortin, M.-J., et al. (2019). Analysing ecological networks of species interactions. *Biological Reviews*, 94, 16–36.
- Eklöf, A. & Ebenman, B. (2006). Species loss and secondary extinctions in simple and complex model communities. *Journal of Animal Ecology*, 75, 239–246.
- Elith, J. & Leathwick, J.R. (2009). Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. *Annual Review of Ecology, Evolution, and Systematics*, 40, 677–697.
- Elton, C.S. (1927). *Animal ecology*. Macmillan Co., New York.
- Engler, R. & Guisan, A. (2009). MigClim: Predicting plant distribution and dispersal in a changing climate. *Diversity and Distributions*, 15, 590–601.
- Funk, J.L., Larson, J.E., Ames, G.M., Butterfield, B.J., Cavender-Bares, J., Firn, J., et al. (2017). Revisiting the Holy Grail: using plant functional traits to understand ecological processes. *Biological Reviews*, 92, 1156–1173.
- Galiana, N., Barros, C., Braga, J., Ficetola, G.F., Maiorano, L., Thuiller, W., et al. (2021). The spatial scaling of food web structure across European biogeographical regions. *Ecography*, 44, 653–664.
- Galiana, N., Lurgi, M., Claramunt-López, B., Fortin, M.-J., Leroux, S., Cazelles, K., et al. (2018). The spatial scaling of species interaction networks. *Nat Ecol Evol*, 2, 782–790.

- Gause, G.F. (1934). Experimental Analysis of Vito Volterra's Mathematical Theory of the Struggle for Existence. *Science*, 79, 16–17.
- Gill, F.B. (2007). *Ornithology*. 3rd ed. W.H. Freeman, New York.
- Godsoe, W., Jankowski, J., Holt, R.D. & Gravel, D. (2017). Integrating Biogeography with Contemporary Niche Theory. *Trends in Ecology & Evolution*, 32, 488–499.
- Gómez, C., Tenorio, E.A., Montoya, P. & Cadena, C.D. (2016). Niche-tracking migrants and niche-switching residents: evolution of climatic niches in New World warblers (Parulidae). *Proceedings of the Royal Society B: Biological Sciences*, 283, 20152458.
- Gómez, J.M., Verdú, M. & Perfectti, F. (2010). Ecological interactions are evolutionarily conserved across the entire tree of life. *Nature*, 465, 918–921.
- Gounand, I., Harvey, E., Little, C.J. & Altermatt, F. (2018). Meta-Ecosystems 2.0: Rooting the Theory into the Field. *Trends in Ecology & Evolution*, 33, 36–46.
- Gravel, D., Baiser, B., Dunne, J.A., Kopalke, J.-P., Martinez, N.D., Nyman, T., et al. (2019). Bringing Elton and Grinnell together: a quantitative framework to represent the biogeography of ecological interaction networks. *Ecography*, 42, 401–415.
- Gravel, D., Canham, C.D., Beaudet, M. & Messier, C. (2006). Reconciling niche and neutrality: the continuum hypothesis. *Ecology Letters*, 9, 399–409.
- Gravel, D., Massol, F., Canard, E., Mouillot, D. & Mouquet, N. (2011). Trophic theory of island biogeography. *Ecology Letters*, 14, 1010–1016.
- Gravel, D., Poisot, T., Albouy, C., Velez, L. & Mouillot, D. (2013). Inferring food web structure from predator-prey body size relationships. *Methods Ecol Evol*, 4, 1083–1090.
- Griffin-Nolan, R.J., Bushey, J.A., Carroll, C.J.W., Challis, A., Chieppa, J., Garbowski, M., et al. (2018). Trait selection and community weighting are key to understanding ecosystem responses to changing precipitation regimes. *Functional Ecology*, 32, 1746–1756.

- Grilli, J., Adorisio, M., Suweis, S., Barabás, G., Banavar, J.R., Allesina, S., *et al.* (2017). Feasibility and coexistence of large ecological communities. *Nat Commun*, 8, 14389.
- Grinnell, J. (1917). The Niche-Relationships of the California Thrasher. *The Auk*, 34, 427–433.
- Guisan, A., Tingley, R., Baumgartner, J.B., Naujokaitis-Lewis, I., Sutcliffe, P.R., Tulloch, A.I.T., *et al.* (2013). Predicting species distributions for conservation decisions. *Ecology Letters*, 16, 1424–1435.
- Guisan, A. & Zimmermann, N.E. (2000). Predictive habitat distribution models in ecology. *Ecological Modelling*, 135, 147–186.
- Harvey, E., Gounand, I., Ward, C.L. & Altermatt, F. (2017). Bridging ecology and conservation: from ecological networks to ecosystem function. *Journal of Applied Ecology*, 54, 371–379.
- Hillebrand, H. (2004). On the Generality of the Latitudinal Diversity Gradient. *The American Naturalist*, 163, 192–211.
- Holt, B.G., Lessard, J.-P., Borregaard, M.K., Fritz, S.A., Araújo, M.B., Dimitrov, D., *et al.* (2013). An Update of Wallace's Zoogeographic Regions of the World. *Science*, 339, 74–78.
- Holt, R.D. (2009). Bringing the Hutchinsonian niche into the 21st century: Ecological and evolutionary perspectives. *PNAS*, 106, 19659–19665.
- Hortal, J., de Bello, F., Diniz-Filho, J.A.F., Lewinsohn, T.M., Lobo, J.M. & Ladle, R.J. (2015). Seven Shortfalls that Beset Large-Scale Knowledge of Biodiversity. *Annu. Rev. Ecol. Evol. Syst.*, 46, 523–549.
- Houlihan, J.E., McKinney, S.T., Anderson, T.M. & McGill, B.J. (2017). The priority of prediction in ecological understanding. *Oikos*, 126, 1–7.
- Hubbell, S.P. (2001). *The unified neutral theory of biodiversity and biogeography*. Monographs in population biology. Princeton University Press, Princeton.

- Humboldt, A. von, Bonpland, A., Jackson, S.T. & Romanowski, S. (2008). *Essay on the geography of plants*. University of Chicago Press, Chicago.
- Hutchinson, G.E. (1957). Concluding Remarks. *Cold Spring Harbor Symposia on Quantitative Biology*, 22, 415–427.
- Hutchinson, G.E. (1961). The Paradox of the Plankton. *The American Naturalist*, 95, 137–145.
- Jordano, P. (2016). Sampling networks of ecological interactions. *Functional Ecology*, 30, 1883–1893.
- Kearney, M., Simpson, S.J., Raubenheimer, D. & Helmuth, B. (2010). Modelling the ecological niche from functional traits. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365, 3469–3483.
- Keddy, P.A. (1992). Assembly and response rules: two goals for predictive community ecology. *Journal of Vegetation Science*, 3, 157–164.
- La Sorte, F.A., Horton, K.G., Johnston, A., Fink, D. & Auer, T. (2022). Seasonal associations with light pollution trends for nocturnally migrating bird populations. *Ecosphere*, 13.
- Laigle, I., Aubin, I., Digel, C., Brose, U., Boulangeat, I. & Gravel, D. (2018). Species traits as drivers of food web structure. *Oikos*, 127, 316–326.
- Lamanna, C., Blonder, B., Violle, C., Kraft, N.J.B., Sandel, B., Šimová, I., et al. (2014). Functional trait space and the latitudinal diversity gradient. *Proceedings of the National Academy of Sciences*, 111, 13745–13750.
- Letten, A.D., Ke, P.-J. & Fukami, T. (2017). Linking modern coexistence theory and contemporary niche theory. *Ecological Monographs*, 87, 161–177.
- Lindeman, R.L. (1942). The Trophic-Dynamic Aspect of Ecology. *Ecology*, 23, 399–417.
- Lundgren, E.J., Bergman, J., Trepel, J., le Roux, E., Monsarrat, S., Kristensen, J.A., et al. (2024). Functional traits—not nativeness—shape the effects of large mammalian herbivores on plant communities. *Science*, 383, 531–537.

- MacArthur, R. (1972). *Geographical Ecology: Patterns in the Distribution of Species*. Princeton University Press, Princeton.
- MacArthur, R.H. & Wilson, E.O. (2001). *The theory of island biogeography*. Princeton landmarks in biology. 13th printing and first Princeton landmarks in biology ed. Princeton university press, Princeton Oxford.
- Marra, P.P., Cohen, E.B., Loss, S.R., Rutter, J.E. & Tonra, C.M. (2015). A call for full annual cycle research in animal ecology. *Biology Letters*, 11, 20150552.
- Matthews, T.J. & Whittaker, R.J. (2015). REVIEW: On the species abundance distribution in applied ecology and biodiversity management. *Journal of Applied Ecology*, 52, 443–454.
- May, R.M. (1972). Will a Large Complex System be Stable? *Nature*, 238, 413–414.
- Mayr, E. (1956). Geographical character gradients and climatic adaptation. *Evolution*, 10, 105–108.
- McDonald-Madden, E., Sabbadin, R., Game, E.T., Baxter, P.W.J., Chadès, I. & Possingham, H.P. (2016). Using food-web theory to conserve ecosystems. *Nat Commun*, 7, 1–8.
- McGill, B.J. (2019). The what, how and why of doing macroecology. *Global Ecology and Biogeography*, 28, 6–17.
- McGill, B.J., Enquist, B.J., Weiher, E. & Westoby, M. (2006). Rebuilding community ecology from functional traits. *Trends in Ecology & Evolution*, 21, 178–185.
- McInerny, G.J. & Etienne, R.S. (2012). Ditch the niche – is the niche a useful concept in ecology or species distribution modelling? *Journal of Biogeography*, 39, 2096–2102.
- McKinnon, L., Smith, P.A., Nol, E., Martin, J.L., Doyle, F.I., Abraham, K.F., et al. (2010). Lower Predation Risk for Migratory Birds at High Latitudes. *Science*, 327, 326–327.
- McMeans, B.C., McCann, K.S., Humphries, M., Rooney, N. & Fisk, A.T. (2015). Food Web Structure in Temporally-Forced Ecosystems. *Trends in Ecology & Evolution*, 30, 662–672.

- Mittelbach, G.G., Schemske, D.W., Cornell, H.V., Allen, A.P., Brown, J.M., Bush, M.B., *et al.* (2007). Evolution and the latitudinal diversity gradient: speciation, extinction and biogeography. *Ecology Letters*, 10, 315–331.
- Morales-Castilla, I., Matias, M.G., Gravel, D. & Araújo, M.B. (2015). Inferring biotic interactions from proxies. *Trends in Ecology & Evolution*, 30, 347–356.
- Nakazawa, Y., Peterson, A.T., Martínez-Meyer, E. & Navarro-Sigüenza, A.G. (2004). SEASONAL NICHES OF NEARCTIC-NEOTROPICAL MIGRATORY BIRDS: IMPLICATIONS FOR THE EVOLUTION OF MIGRATION. *Auk*, 121, 610.
- Norris, D.R., Marra, P.P., Kyser, T.K., Sherry, T.W. & Ratcliffe, L.M. (2004). Tropical winter habitat limits reproductive success on the temperate breeding grounds in a migratory bird. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 271, 59–64.
- O'Connor, L.M.J., Pollock, L.J., Braga, J., Ficetola, G.F., Maiorano, L., Martinez-Almoyna, C., *et al.* (2020). Unveiling the food webs of tetrapods across Europe through the prism of the Eltonian niche. *Journal of Biogeography*, 47, 181–192.
- Pacifici, M., Visconti, P., Butchart, S.H.M., Watson, J.E.M., Cassola, F.M. & Rondinini, C. (2017). Species' traits influenced their response to recent climate change. *Nature Clim Change*, 7, 205–208.
- Pellissier, L., Albouy, C., Bascompte, J., Farwig, N., Graham, C., Loreau, M., *et al.* (2018). Comparing species interaction networks along environmental gradients. *Biological Reviews*, 93, 785–800.
- Pianka, E.R. (1966). Latitudinal Gradients in Species Diversity: A Review of Concepts. *The American Naturalist*, 100, 33–46.
- Pichler, M., Boreux, V., Klein, A.-M., Schleuning, M. & Hartig, F. (2020). Machine learning algorithms to infer trait-matching and predict species interactions in ecological networks. *Methods in Ecology and Evolution*, 11, 281–293.

Pimm, S.L., Lawton, J.H. & Cohen, J.E. (1991). Food web patterns and their consequences. *Nature*, 350, 669–674.

Poisot, T., Bergeron, G., Cazelles, K., Dallas, T., Gravel, D., MacDonald, A., *et al.* (2021). Global knowledge gaps in species interaction networks data. *J Biogeogr*, jbi.14127.

Poisot, T., Canard, E., Mouillot, D., Mouquet, N. & Gravel, D. (2012). The dissimilarity of species interaction networks. *Ecology Letters*, 15, 1353–1361.

Poisot, T., Mouquet, N. & Gravel, D. (2013). Trophic complementarity drives the biodiversity–ecosystem functioning relationship in food webs. *Ecology Letters*, 16, 853–861.

Poisot, T., Stouffer, D.B. & Gravel, D. (2015). Beyond species: why ecological interaction networks vary through space and time. *Oikos*, 124, 243–251.

Pollock, L.J., Morris, W.K. & Vesk, P.A. (2012). The role of functional traits in species distributions revealed through a hierarchical model. *Ecography*, 35, 716–725.

Pollock, L.J., O'Connor, L.M.J., Mokany, K., Rosauer, D.F., Talluto, M.V. & Thuiller, W. (2020). Protecting Biodiversity (in All Its Complexity): New Models and Methods. *Trends in Ecology & Evolution*, 35, 1119–1128.

Pollock, L.J., Tingley, R., Morris, W.K., Golding, N., O'Hara, R.B., Parris, K.M., *et al.* (2014). Understanding co-occurrence by modelling species simultaneously with a Joint Species Distribution Model (JSDM). *Methods in Ecology and Evolution*, 5, 397–406.

Pomeranz, J.P.F., Thompson, R.M., Poisot, T. & Harding, J.S. (2019). Inferring predator–prey interactions in food webs. *Methods in Ecology and Evolution*, 10, 356–367.

Pontarp, M., Bunnefeld, L., Cabral, J.S., Etienne, R.S., Fritz, S.A., Gillespie, R., *et al.* (2019). The Latitudinal Diversity Gradient: Novel Understanding through Mechanistic Eco-evolutionary Models. *Trends in Ecology & Evolution*, 34, 211–223.

- Ponti, R. & Sannolo, M. (2023). The importance of including phenology when modelling species ecological niche. *Ecography*, 2023, e06143.
- Preston, F.W. (1960). Time and Space and the Variation of Species. *Ecology*, 41, 612–627.
- Pringle, R.M. & Hutchinson, M.C. (2020). Resolving Food-Web Structure. *Annual Review of Ecology, Evolution and Systematics*, 51, 55–80.
- Pulliam, H. r. (2000). On the relationship between niche and distribution. *Ecology Letters*, 3, 349–361.
- Reich, P.B., Wright, I.J., Cavender-Bares, J., Craine, J.M., Oleksyn, J., Westoby, M., et al. (2003). The Evolution of Plant Functional Variation: Traits, Spectra, and Strategies. *International Journal of Plant Sciences*, 164, S143–S164.
- Reudink, M.W., Marra, P.P., Kyser, T.K., Boag, P.T., Langin, K.M. & Ratcliffe, L.M. (2009). Non-breeding season events influence sexual selection in a long-distance migratory bird. *Proceedings of the Royal Society B: Biological Sciences*, 276, 1619–1626.
- Rooney, N., McCann, K.S. & Moore, J.C. (2008). A landscape theory for food web architecture. *Ecology Letters*, 11, 867–881.
- Rosenberg, K.V., Dokter, A.M., Blancher, P.J., Sauer, J.R., Smith, A.C., Smith, P.A., et al. (2019). Decline of the North American avifauna. *Science*, 366, 120–124.
- Rossberg, A.G., Bränström, Å. & Dieckmann, U. (2010). How trophic interaction strength depends on traits. *Theor Ecol*, 3, 13–24.
- Runge, C.A., Watson, J.E.M., Butchart, S.H.M., Hanson, J.O., Possingham, H.P. & Fuller, R.A. (2015). Protected areas and global conservation of migratory birds. *Science*, 350, 1255–1258.
- Saravia, L.A., Marina, T.I., Kristensen, N.P., De Troch, M. & Momo, F.R. (2022). Ecological network assembly: How the regional metaweb influences local food webs. *Journal of Animal Ecology*, 1365-2656.13652.

- Shipley, B.R., Bach, R., Do, Y., Strathearn, H., McGuire, J.L. & Dilkina, B. (2022). megaSDM: integrating dispersal and time-step analyses into species distribution models. *Ecography*, 2022.
- Soberón, J. (2007). Grinnellian and Eltonian niches and geographic distributions of species. *Ecology Letters*, 10, 1115–1123.
- Somveille, M., Manica, A. & Rodrigues, A.S.L. (2019). Where the wild birds go: explaining the differences in migratory destinations across terrestrial bird species. *Ecography*, 42, 225–236.
- Somveille, M., Rodrigues, A.S.L. & Manica, A. (2015). Why do birds migrate? A macroecological perspective. *Global Ecology and Biogeography*, 24, 664–674.
- Staniczenko, P.P.A., Sivasubramaniam, P., Suttle, K.B. & Pearson, R.G. (2017). Linking macroecology and community ecology: refining predictions of species distributions using biotic interaction networks. *Ecology Letters*, 20, 693–707.
- Stouffer, D.B., Sales-Pardo, M., Sirer, M.I. & Bascompte, J. (2012). Evolutionary Conservation of Species' Roles in Food Webs. *Science*, 335, 1489–1492.
- Strydom, T., Catchen, M.D., Banville, F., Caron, D., Dansereau, G., Desjardins-Proulx, P., et al. (2021). A roadmap towards predicting species interaction networks (across space and time). *Philos Trans R Soc Lond B Biol Sci*, 376, 20210063.
- Studds, C.E., Kendall, B.E., Murray, N.J., Wilson, H.B., Rogers, D.I., Clemens, R.S., et al. (2017). Rapid population decline in migratory shorebirds relying on Yellow Sea tidal mudflats as stopover sites. *Nat Commun*, 8, 14895.
- Thompson, R.M., Brose, U., Dunne, J.A., Hall, R.O., Hladyz, S., Kitching, R.L., et al. (2012). Food webs: reconciling the structure and function of biodiversity. *Trends in Ecology & Evolution*, 27, 689–697.
- Thompson, R.M. & Townsend, C.R. (2005). Energy Availability, Spatial Heterogeneity and Ecosystem Size Predict Food-Web Structure in Streams. *Oikos*, 108, 137–148.

- Thuiller, W., Münkemüller, T., Lavergne, S., Mouillot, D., Mouquet, N., Schiffers, K., *et al.* (2013). A road map for integrating eco-evolutionary processes into biodiversity models. *Ecology Letters*, 16, 94–105.
- Thurman, L.L., Barner, A.K., Garcia, T.S. & Chestnut, T. (2019). Testing the link between species interactions and species co-occurrence in a trophic network. *Ecography*, 42, 1658–1670.
- Tingley, M.W., Monahan, W.B., Beissinger, S.R. & Moritz, C. (2009). Birds track their Grinnellian niche through a century of climate change. *Proceedings of the National Academy of Sciences*, 106, 19637–19643.
- Tylianakis, J.M. & Morris, R.J. (2017). Ecological Networks Across Environmental Gradients. *Annu. Rev. Ecol. Evol. Syst.*, 48, 25–48.
- Urban, M.C., Bocedi, G., Hendry, A.P., Mihoub, J.-B., Peer, G., Singer, A., *et al.* (2016). Improving the forecast for biodiversity under climate change. *Science*, 353, aad8466–aad8466.
- Vandewalle, M., de Bello, F., Berg, M.P., Bolger, T., Dolédec, S., Dubs, F., *et al.* (2010). Functional traits as indicators of biodiversity response to land use changes across ecosystems and organisms. *Biodivers Conserv*, 19, 2921–2947.
- Vellend, M. (2010). Conceptual Synthesis in Community Ecology. *The Quarterly Review of Biology*, 85, 183–206.
- Vesk, P.A., Morris, W.K., Neal, W.C., Mokany, K. & Pollock, L.J. (2021). Transferability of trait-based species distribution models. *Ecography*, 44, 134–147.
- Violle, C., Navas, M.-L., Vile, D., Kazakou, E., Fortunel, C., Hummel, I., *et al.* (2007). Let the concept of trait be functional! *Oikos*, 116, 882–892.
- Volterra, V. (1928). Variations and Fluctuations of the Number of Individuals in Animal Species living together. *ICES Journal of Marine Science*, 3, 3–51.

- Wallace, A.R. (2011). *The Geographical Distribution of Animals: With a Study of the Relations of Living and Extinct Faunas as Elucidating the Past Changes of the Earth's Surface*. 1st edn. Cambridge University Press.
- Watanabe, Y.Y. (2016). Flight mode affects allometry of migration range in birds. *Ecology Letters*, 19, 907–914.
- Whittaker, R.J., Willis, K.J. & Field, R. (2001). Scale and species richness: towards a general, hierarchical theory of species diversity. *Journal of Biogeography*, 28, 453–470.
- Wiens, J.J. (2011). The niche, biogeography and species interactions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366, 2336–2350.
- Wilcove, D.S. & Wikelski, M. (2008). Going, Going, Gone: Is Animal Migration Disappearing. *PLOS Biology*, 6, e188.
- Williams, R.J. & Purves, D.W. (2011). The probabilistic niche model reveals substantial variation in the niche structure of empirical food webs. *Ecology*, 92, 1849–1857.
- Windsor, F.M., van den Hoogen, J., Crowther, T.W. & Evans, D.M. (2023). Using ecological networks to answer questions in global biogeography and ecology. *Journal of Biogeography*, 50, 57–69.
- Winger, B.M., Auteri, G.G., Pegan, T.M. & Weeks, B.C. (2019). A long winter for the Red Queen: rethinking the evolution of seasonal migration. *Biological Reviews*, 94, 737–752.
- Wisz, M.S., Pottier, J., Kissling, W.D., Pellissier, L., Lenoir, J., Damgaard, C.F., et al. (2013). The role of biotic interactions in shaping distributions and realised assemblages of species: implications for species distribution modelling. *Biological Reviews*, 88, 15–30.
- Xu, Y., Si, Y., Takekawa, J., Liu, Q., Prins, H.H.T., Yin, S., et al. (2020). A network approach to prioritize conservation efforts for migratory birds. *Conservation Biology*, 34, 416–426.
- Zakharova, L., Meyer, K.M. & Seifan, M. (2019). Trait-based modelling in ecology: A review of two decades of research. *Ecological Modelling*, 407, 108703.

- Zhang, W., Wei, J. & Xu, Y. (2023). Prioritizing global conservation of migratory birds over their migration network. *One Earth*, 6, 1340–1349.
- Zurell, D., Gallien, L., Graham, C.H. & Zimmermann, N.E. (2018). Do long-distance migratory birds track their niche through seasons? *Journal of Biogeography*, 45, 1459–1468.
- Zurell, D., Jeltsch, F., Dormann, C.F. & Schröder, B. (2009). Static species distribution models in dynamically changing systems: how good can predictions really be? *Ecography*, 32, 733–744.

# **Chapter 1: Addressing the Eltonian shortfall with trait-based interaction models**

**Dominique Caron<sup>1,2</sup>, Luigi Maiorano<sup>3</sup>, Wilfried Thuiller<sup>4</sup>, Laura J. Pollock<sup>1,2</sup>**

## **Affiliations:**

<sup>1</sup> Department of Biology, McGill University, Montreal, QC, Canada

<sup>2</sup> Quebec Centre for Biodiversity Sciences, Montreal, QC, Canada

<sup>3</sup> Department of Biology and Biotechnologies "Charles Darwin", Sapienza University of Rome, viale dell'Università 32, 00185 Rome, Italy

<sup>4</sup> Univ. Grenoble Alpes, Univ. Savoie Mont Blanc, CNRS, LECA, Laboratoire d'Ecologie Alpine, F-38000 Grenoble, France.

The following has been published in *Ecology Letters* under the reference below:

Caron, D., Maiorano, L., Thuiller, W., & Pollock, L. J. (2022). Addressing the Eltonian shortfall with trait-based interaction models. *Ecology Letters*, 25(4), 889-899. <https://doi.org/10.1111/ele.13966>

Reproduced here under the terms of the Creative Commons Attribution License.

## **Abstract**

We have very limited knowledge of how species interact in most communities and ecosystems despite trophic relationships being fundamental for linking biodiversity to ecosystem functioning. A promising approach to fill this gap is to predict interactions based on functional traits, but many questions remain about how well we can predict interactions for different taxa, ecosystems, and amounts of input data. Here, we built a new traits-based model of trophic interactions for European vertebrates and found that even models calibrated with 0.1% of the interactions (100 out of 71k) estimated the full European vertebrate food web reasonably well. However, predators were easier to predict than prey, especially for some clades (e.g. fowl and storks) and local food web connectance was consistently overestimated. Our results demonstrate the ability to rapidly generate food webs when empirical data are lacking - an important step towards a more complete and spatially-explicit description of food webs.

## Introduction

Food webs are collections of trophic interactions describing both the composition and structure of communities. Knowing the trophic structure of a food web is critical to understand how energy flows through ecosystems (Thompson *et al.* 2012), how populations respond to perturbations through time (Zhao *et al.* 2019), and how species are distributed in space (Wisz *et al.* 2013). The food web is also critical for conservation efforts (Harvey *et al.* 2017; Pollock *et al.* 2020) and can be an indicator of the threat of extinction (McDonald-Madden *et al.* 2016). If food web links are lost more quickly than species, then food webs can become simplified and homogenous with fewer trophic levels (Laliberté & Tylianakis 2010; Estes *et al.* 2011; Valiente-Banuet *et al.* 2015). Despite this central importance of food webs in ecology and conservation, we have yet to fully understand predator-prey relationships for many species even in relatively well-studied areas.

One of the main reasons for this lack of understanding is the Eltonian shortfall – “*the lack of knowledge about interactions among species or among groups of species*” (Hortal *et al.* 2015). Available data on food webs are mostly restricted to the United States and Europe similar to other types of species interaction networks (Hortal *et al.* 2015; Cameron *et al.* 2019; Poisot *et al.* 2021). Compounding the problem even further, the sampling effort to detect interactions far exceeds the effort needed to detect species because they require the simultaneous detection of both interacting species and the interaction (Chacoff *et al.* 2012; Jordano 2016). Consequently, rare interactions are often missed while those of dominant species are overestimated. So, even when available, food web datasets are often incomplete and biased. A promising solution is to fill gaps in empirical food webs with expert knowledge and literature review (e.g., Piechnik *et al.* 2008; Maiorano *et al.* 2020), but this approach remains limited to well-studied systems.

Despite the size of the problem, there is reason to be optimistic about the potential for predicting species interactions to fill gaps in food web data. In neutral food webs, the abundances of coexisting species predict their probability of interacting (Canard *et al.* 2012). In many cases, trophic interactions can be better predicted if they conform to a predictable set of phylogenetic and functional traits (Morales-Castilla *et al.* 2015). Niche theory predicts that two species interact if the foraging traits of the predator match the vulnerability traits of the prey (Williams & Martinez 2000; Gravel *et al.* 2016). This trait-matching framework serves as the basis for most studies aiming to predict interactions (e.g., Gravel *et al.* 2013; Bartomeus *et al.* 2016; Pichler *et al.* 2020), which have shown promising results. Notably, Eklöf *et al.*, (2013) found that 3-5 traits suffice to predict most trophic interactions, whereas Gravel *et al.*, (2013) showed that predictions of trophic interactions in marine ecosystems are robust to sampling effort. These findings suggest that we can potentially predict missing trophic interactions even when data are sparse.

Yet, there have been few large-scale, multi-clade tests of predictive models of trophic interactions. Many studies have used trait-matching models in systems where the relationships between the traits of predators and their prey are expected to be strong, such as marine food webs (e.g., Gravel *et al.* 2013; Laigle *et al.* 2018; Albouy *et al.* 2019). Other studies have built more complex models which can outperform simpler models (e.g., Rohr *et al.* 2016), but have many parameters, relying on large datasets that are often unavailable. Therefore, models based on relatively simple trait-matching relationships are likely the best option for a wide range of taxa that have available trait data, but are they realistic enough to make good predictions across a diverse set of taxa and ecosystems? We need a better understanding of the amount of data needed to make reliable predictions and how general the matching rules are across clades and space.

Here, we built a model to predict trophic interactions based on functional traits that are widely available. We designed our model to be a very general model that can be applied across vertebrate groups using a Bayesian linear model relating trophic interaction (and non-interactions) to a relatively small set of predictor variables. These predictors describe the foraging ability of the predator, the vulnerability of the prey, and the trait-match of interacting species. The relatively simplistic model structure allows for fitting the model with few training data. We test our model using both the recently assembled food web of all terrestrial vertebrates in Europe (Maiorano *et al.* 2020) and the Global Biotic Interactions (GloBI) database (Poelen *et al.* 2014), and use the model to determine: (1) how much data are needed to accurately predict the entire European food web, (2) which traits make the best predictions, (3) how input data (empirical versus multi-sourced data) influence predictions, (4) which taxa are easier or harder to make predictions for, and (5) how well the properties of local food webs can be predicted across space.

## Materials and methods

### *Study area, species, and functional traits*

Our study focuses on trophic interactions among all terrestrial vertebrates of Europe: mammals, breeding birds, reptiles, and amphibians. For each species, we extracted five functional traits from Thuiller *et al.* (2015): diet, nesting habitat, activity time, foraging behavior, and body mass (Appendix S1). Because body mass was missing for many amphibians - 52% missing for frogs (*Anura*) and 46% for salamanders (*Caudata*) - we imputed missing body mass from available body length information for these two groups separately. These imputations were justified by the strong relationship between body mass and length in our study ( $r^2 = 0.75$  and 0.84 for frogs and salamanders respectively; Appendix S2).

and others (Deichmann *et al.* 2008). In all, we gathered these five functional traits for 1055 species: 101 amphibians, 507 birds, 267 mammals, and 180 reptiles.

Using diet, nesting habitat, activity time, foraging behavior, and body mass, we calculated seven predictor variables for each species pair (Table 1). These predictors of predator-prey interactions can be categorized into three types: foraging traits, vulnerability traits, and matching traits (Gravel *et al.* 2016; Rohr *et al.* 2016). Foraging traits (body mass of the predator and its foraging behavior) influence the number of prey of a given predator, whereas vulnerability traits (body mass of the prey) influence the number of predators of a given prey. Matching traits influence the feasibility of the interaction (difference in body mass and diet match) or the encounter rate of species (activity time and habitat match).

#### *Interaction data*

We extracted interaction data from the trophic metaweb of European terrestrial vertebrates (referred to as the Metaweb; Maiorano *et al.*, 2020). A metaweb documents all potential interactions between all species at the regional scale (Dunne 2006). Maiorano *et al.* (2020) compiled all potential trophic interactions and non-interactions between all terrestrial vertebrates of Europe from guide books, published papers, and completed by expert opinion. As all pairs of species were assessed as to whether they could potentially interact, we assumed all zeros are true non-interactions rather than resulting from missing data. We extracted the interactions and non-interactions between all of the 1 055 species for which we had functional traits. This represents 71 417 potential interactions and 1 041 608 non-interactions.

Because the Metaweb documents *potential* trophic interactions, we also used a database of empirical observations of species interactions. We extracted all interactions reported in the Global Biotic Interactions (GloBI) platform (Poelen *et al.* 2014). The Metaweb and GloBI are very different in terms of data coverage. While the Metaweb documents only trophic interactions in Europe, GloBI aggregates trophic (and non-trophic) interactions from anywhere in the world. The absences of interactions in the Metaweb can be interpreted as likely to be true absences (i.e., low false negative rate), whereas absences of interactions in GloBI cannot (i.e., high false negative rate). Conversely, the presences of interactions in the Metaweb are more uncertain (i.e., high false positive rate) than in GloBI. We used the package rglobi of the R software to extract all trophic interactions between any of the focal 1 055 species. In all, we extracted 291 trophic interactions from GloBI involving 194 different species (75 predator species and 146 prey species).

### *Predictive model*

We modelled the occurrence of a food web interaction for each pair of species as a function of their traits using Bayesian generalized linear models (GLM). We assumed that the occurrence of interaction between species  $i$  and  $j$ ,  $L_{ij}$ , is Bernouilli distributed. The corresponding probability of interaction was modelled as the inverse *logit* of a linear function with a common intercept,  $\alpha$ , and a set of linear coefficients  $\beta$  associated with the seven predictors  $T_{ij}$  (Table 1):

$$\text{logit}(P(L = 1)) = \alpha + \sum_{k=1}^7 \beta_k T_{ijk} \quad (\text{Eq. 1})$$

We measured the predictive performance of each model on independent validation datasets using the area under the receiver operating characteristic curve (AUC). AUC varies from 0.5 to 1 where 0.5 indicates that the model failed to rank interactions higher than non-interactions (i.e., random prediction), and 1 indicates that the model systematically ranked interactions higher than non-

interactions (i.e., perfect prediction). We also measured the area under the precision-recall curve, the true positive rate, true negative rate, and true-skill statistics to see if our results were robust to the choice of performance metric (Appendix S8).

Before fitting the GLMs, we scaled each continuous predictor by subtracting it by its mean and dividing by two times its standard deviation, so that the coefficients of the scaled continuous predictors are directly comparable to coefficients of unscaled binary predictors (Gelman 2008). We used Bayesian inference and Markov chain Monte Carlo algorithm to estimate the model parameters ( $\alpha$  and  $\beta_k$ ). We used weakly informative priors for the parameters:

$$\alpha \sim Normal(mean = 0, sd = 10) \quad (\text{Eq. 2})$$

$$\beta_k \sim Normal(mean = 0, sd = 1) \quad (\text{Eq. 3})$$

where  $sd$  is the standard deviation of the prior distribution. We ran 3 chains, each with 1000 warm-up iterations, followed by 5 000 iterations for inference. We diagnosed convergence visually of a few test runs, and calculated the potential scale reduction factor,  $\hat{R}$ , for all runs (Gelman & Rubin 1992; Appendix S3). We conducted the Bayesian analyses using the package *greta* in R (Golding 2019).

#### *Predicting trophic interactions with models trained on the Metaweb*

We used the model described above to make predictions for predator-prey pairs in the Metaweb, and determine the information needed to make reliable predictions. More specifically, we determined: (1) the number of trophic interactions needed to calibrate models that could recover most of the Metaweb, and (2) which of the predictors made the most important contribution. To determine the information needed to make predictions, we trained a set of 75 models with progressively more trophic interactions from the Metaweb going from 2 to 5000 pairwise interactions, with 10 000 non-interactions. Each of

the 75 models used the same set of seven predictors (Table 1). We then measured how well these models predicted the Metaweb by comparing the predictions generated by the model to an independent validation dataset, which was a sample of 1% of the Metaweb (Metaweb validation dataset; random sampling had little effect on predictive performance; Appendix S4). To compare the importance of trait-based predictor variables, we compared the mean coefficient values of all scaled predictors (Gelman 2008) from a model calibrated on the entire Metaweb.

We also measured how well each of the models described above predicts the empirical interactions in GloBi. To do so, we compared predictions generated by each model to the 291 trophic interactions we extracted from GloBi. Because GloBI only includes interactions, we added 3 845 pairs of species that do not interact in the Metaweb (non-interactions) to the GloBI validation dataset. We added 3 845 non-interactions to make the prevalence of interactions in the GloBI validation dataset comparable to the prevalence in the Metaweb validation dataset.

#### *Predicting trophic interactions with models trained on GloBI*

We were also interested in how well we could predict the entire Metaweb without using any information from the Metaweb itself. To do this, we combined information on species traits with the observed interactions included in GloBI. We fitted 40 predictive models using a progressively increasing number of trophic interactions extracted from GloBI: from 2 to 291 trophic interactions combined with 10 000 pseudo-absences. To extract pseudo-absences, we randomly drew pairs of species for which no interaction is documented in GloBI. We used pseudo-absences to avoid using any information from the Metaweb. We measured performance of these models with respect to the Metaweb validation dataset, and to the GloBI validation dataset.

### *Predictions of interactions for different taxa*

In addition to the overall predictive performance of the models, we wanted to identify whether interactions between certain clades and different types of species (e.g., specialists versus generalists) were more or less predictable with our model. To do this, we first trained a model (*master* model) with 10% of the European Metaweb (7 157 trophic interactions and 111 287 non-interactions) and predicted all trophic interactions and non-interactions in the entire food web. We measured the performance of the model to predict all interactions (i.e., the prey and predators) of every species. We further compared these performances for individual species to explore whether interactions of specialists (i.e., species interacting with few species) were harder to predict than interactions of generalists (i.e., species interacting with many species). Specifically, we modelled the performance for each species (the AUC value) as a function of the species generality using three bayesian generalized linear mixed models (GLMM<sub>all</sub>, GLMM<sub>prey</sub>, GLMM<sub>predator</sub>) with random intercepts and slopes for the species group (i.e., order), and a logit-link function. All species were included in each model, but different representations of generality of each species (GLMM<sub>all</sub>: the number of total interactions; GLMM<sub>prey</sub>: the number of prey; GLMM<sub>predator</sub>: the number of predators) were included as fixed effects in the three separate models. The fixed effects were log-transformed and scaled before running the GLMMs. We fitted the GLMMs using the package *greta* in R (details in appendix S9).

Finally, we performed a miscalibration analysis to investigate the ecological differences making the trophic interactions of some groups less predictable than others. We first trained group-specific models for each order of predators. For example, considering frogs (*Anura*), we trained a predictive model only using interactions (and non-interactions) involving a frog species as predator. For many orders, most species shared very similar foraging behaviors. This caused some group-specific models not to converge or to over-fit the data. For this reason, we excluded foraging behaviors from the predictors in

the miscalibration analysis. We compared the group-specific models to a *general* model trained on the entire Metaweb (71 417 trophic interactions and 1 113 025 non-interactions). Specifically, we calculated the differences between the parameters inferred for the *general* model to the parameters inferred for the group-specific models to measure the miscalibration of the *general* model for each group. By doing so, we identified the miscalibrated coefficients causing the general model to incorrectly predict the prey of specific predator groups. In contrast to the *master* model, the *general* model did not include foraging behaviors as predictors to make it comparable to group-specific models.

#### *Predictions of interactions across space*

In addition to model performance for different types of species, we also tested how well the model performed in different regions and for local assemblages. This tests the generality of how traits can predict potential trophic interactions with a range of ecological constraints and species pools. First, we used the *master* model described above to make predictions for ‘local’ food webs based on distribution data for each 10km pixel across Europe. Second, we addressed model transferability by training models in one region and using that model to predict to another region. If a model trained in the arctic, for example, can predict interactions in the Mediterranean with regional specialist species, then this is evidence for the generality of trait-interaction relationships and the ability of these models to extrapolate.

We determined the local/regional webs based on species composition data from Maiorano et al. (2013), where species distributions are determined by the presence of primary habitat within the known species distribution range and validated by field data. Species habitat requirements are determined by experts based on land cover, elevation and distance to water. Species distribution ranges are extracted from atlases (see Maiorano *et al.* 2013 for the full list of atlases). The original data are 300-m cell range

maps with three levels: unsuitable, secondary and primary habitat. We upscaled the distributions data to a 10km pixel equal-area grid, and considered a species present in a cell if it had at least one occurrence of primary habitat.

We created ‘local’ food webs for each 10km pixel across Europe by pruning the Metaweb in every pixel based on the species present in that pixel. We assumed that spatial variation in potential food webs are simply due to species turnover. We are not taking into account intraspecific trait variation, change in abundances or interaction turnovers, which are likely important in the local realization of interactions. We predicted the interactions for each local species assemblage by pruning the predictions of the entire Metaweb (using the *master* model) based on the species pool in each pixel and compared these to the ‘true’ interactions according to the pruned metaweb. If environmental gradients influence the ecological constraints driving trophic interactions, we expect to detect a gradient in the predictability of local food webs. We specifically wanted to investigate whether simpler, less connected food webs (lower connectance) were easier to predict (evaluated with AUC, true positive and true negative rates) than more complex, highly connected food webs.

We created regional food webs by extracting the trophic interactions and non-interactions from the Metaweb between species for which the range intersected in at least 1 % of the European bioregions (see O’Connor et al. 2020) and trained models for each bioregion individually. To investigate the transferability of predictive models across bioregions, we compared the importance of model predictors in each bioregion (e.g. the linear coefficient associated with predator body mass in the arctic versus in the mediterranean bioregion), and how well they could predict the food webs of other bioregions. If environmental gradients influence the ecological constraints driving trophic interactions, we expect poor transferability of models between bioregions.

## Results

*How much data are needed to accurately predict the entire European food web?*

Most pairwise interactions in the Metaweb were predicted reasonably well by models trained on few interactions. Predictive performance increased when more interactions were used to train the model, but performance stabilized around AUC=0.92 at 100 training interactions (0.14% of the total number of interactions in the Metaweb; Fig 1a). Even with as low as 10 interactions (0.014% of all interactions in the Metaweb; Fig 1a), the AUC was over 0.90. Training the model on a lot more presences and absences did not improve substantially model performance. The *master* model, which we trained on 10% of the entire Metaweb to predict the interactions of every species, resulted in an AUC of 0.92, well above the 0.5 expected from a null model.

*Which traits best predict interactions?*

All of the traits used as predictor variables in the models were important for predicting interactions: all corresponding linear coefficients were different from 0 (i.e., 95% credible interval did not include 0). The most important predictors were the match between the diet of the predator and the type of prey (median = 2.29, 95%CrI = [2.27, 2.32]), the body mass of the predator (median = 2.08, 95%CrI = [2.04, 2.12]), the body mass of the prey (median = -2.00, 95%CrI = [-2.03, -1.96]), and the match between the body mass of the predator and the body mass of the prey (median = -1.75, 95%CrI = [-1.79, -1.70]). All parameter estimates with their 95% credible interval are available in Appendix S5.

*How input data (empirical versus multi-sourced data) influences predictions?*

We found that results from models trained on the European Metaweb were consistent with models trained on interactions from GloBI. Models fitted using all GloBI interactions (291 interactions) predicted the Metaweb well ( $AUC=0.91$ ; Fig 1a). Also, all models performed similarly in predicting the realized interactions from GloBI compared to the potential interactions of the Metaweb (Fig 1b).

#### *Which taxa are easier or harder to make predictions for?*

Overall, the *master* model performed well for all groups ( $AUC > 0.75$  for all groups; Fig 2). The variation between groups was mostly due to variation in the ability of the model to predict prey (Fig 2). For example, prey of carnivorans (*Carnivora*), fowl (*Galliformes* and *Anseriformes*), pelicans (*Pelicaniformes*), storks (*Ciconiformes*), and birds of prey (*Falconiformes*, *Charadriiformes*, and *Strigiformes*) were harder to predict on average (Fig 2). In contrast, predicting predators was similarly easy across all prey groups ( $AUC > 0.9$ ). One exception to this pattern was amphibians: the prey of *Caudata* and *Anura* were easier to predict ( $AUC = 0.96$  and  $0.95$  respectively) than their predators ( $AUC = 0.85$  and  $0.78$  respectively).

Interactions of specialists tended to be more predictable than interactions of generalists (Fig 3). The order of the species and the number of interactions explained 56% (bayesian  $R^2_{GLMMall}$ ) of the variation in the logit-AUC (fixed effect estimate =  $-0.49$ ; 95%CrI =  $[-0.54, -0.43]$ ). In general, it is easier to predict species' predators than their prey (Fig 3). Prey of generalist predators tended to be harder to predict ( $GLMM_{prey}$  fixed effect estimate =  $-0.30$ ; 95%CrI =  $[-0.38, -0.22]$ ). Similarly, predators of high vulnerability prey were harder to predict ( $GLMM_{predator}$  fixed effect estimate =  $-0.24$ ; 95%CrI =  $[-0.28, -0.20]$ ).

For most groups in which potential prey were difficult to predict, the group-specific model (i.e. model calibrated on interactions of one predator order) performed much better, with the AUC increasing by 0.1 or more (Appendix S6). Our models highlight how some bird groups (e.g., *Galliformes*, *Anseriformes*, *Pelicaniformes*) tend to feed on prey active at different times during the day and nest in different habitats than the focal predator. This contrasts with woodpeckers (*Piciformes*) for which the matches of activity times and nesting habitats of the prey and the predator are especially important. For other groups (e.g., *Rodentia*, *Eulipotyphla*, *Strigiformes*, *Passeriformes*) the relation between the body mass of the predator and the prey is more important to explain interactions than average. We also found that, for carnivorans (*Carnivora*), the predictive performance of the group-specific model remained low (AUC = 0.63). The coefficients of the different predictors were mostly lower (closer to 0) than the *general* model, suggesting that the interactions of carnivorans are difficult to predict from our set of traits. The complete results of the miscalibration analysis are available in the Appendix S6.

#### *How well the properties of local food webs can be predicted across space?*

Overall, local food webs with more trophic interactions were less predictable than simpler ones (Fig 4). This trend resulted in a slight increase in the predictability of local food webs at higher latitudes (Fig 4). True positive rates were higher than true negative rates across the entire continent which caused the predicted connectance (proportion of possible links that are realized) to be systematically overestimated by 2 to 4 times (Fig 4). All bioregional food webs were similarly predictable ( $0.89 < \text{AUC} < 0.92$ ) by the *master* model and models transferred well from one bioregion to another (Appendix S7). Similarly, trait parameters of each bioregional predictive model did not diverge greatly from the *master* model (Appendix S7).

## Discussion

In this study, we recovered the entire European food web of tetrapods from a fraction of the entire food web and widely available functional traits. Our results indicate that these traits (describing the foraging of predators, vulnerability of prey, and the trait-match between predator and prey) are indeed predictive of the predator-prey pairs that make up large food webs. These trait-interaction relationships appear to be general given: (1) the stability of modeled effects (e.g., the positive effect of body mass differences) across most vertebrate orders and highly contrasting ecosystems, (2) the consistency of predictive power between the complete, yet potential interaction data of the European metaweb and the sparse, empirical data from GloBI, and (3) the efficiency of the model to make good predictions with few input interactions (recovering ~71k interactions with <100 known interactions with an AUC > 0.90). While predictions were mostly reliable, there were exceptions including: poor predictions for the prey of some predator orders, the trophic interactions of generalists and the tendency to overestimate connectance especially in complex food webs, which we discuss further below.

Overall, this generality in trait-interaction relationships and the ability of the model to extrapolate suggests these models could be applied to many ecosystems, even those with very sparse and biased datasets. GloBi only documented 291 trophic interactions compared to the 71 417 trophic interactions documented in the European Metaweb, which illustrate the sparsity of trophic interactions data even in one of the best-studied continents. With very limited data, it is important that models can extrapolate to new conditions (Roberts *et al.* 2017; Santini *et al.* 2021). Our models produced similar results (e.g., had similar coefficients for trait variables) when fitted to different bioregions, and predictive performance of bioregional models did not decrease when one bioregion was used to predict to another. This indicates potential transfer of information from data-rich bioregions to data-poor bioregions. While

traits have been shown to promote transferring information from one region to the next for species distributions (Vesk *et al.* 2021), here we show how trait-based models can similarly transfer information on *potential* interactions to entirely new areas. We need more studies on the factors influencing model transferability in food web interactions to fully appreciate the potential of trait-based models to predict interactions in data-poor systems or under future conditions (Yates *et al.* 2018). The prediction and transfer of information on *realized* interactions is likely to be much harder since they are influenced by local abundances, intraspecific trait variation, and indirect interactions from other species (Poisot *et al.* 2015; Pellissier *et al.* 2018).

While the generality and overall performance of these models are promising, we highlighted some systematic biases. For example, our predictive model systematically overestimated the number of links in local food webs (Fig 4). Interactions were better predicted (true positive rate  $\approx 0.9$ ) than non-interactions (true negative rate  $\approx 0.8$ ), meaning that predicted interactions includes most realized interactions, but also many incorrectly predicted non-interactions (false discovery rate  $\approx 0.7$ ; Appendix S8). Thus, our predictive model should be viewed as a first step toward a correct description of a regional and local food webs by reducing the millions of possible interactions to thousands of feasible ones. These predicted interactions could be used to inform targeted sampling or expert elicitation to get a more accurate picture of the true food web.

In addition, not all trophic interactions were equivalent. The prey of some taxonomic groups were harder to predict than others using a general model (Fig 2). Our miscalibration analysis helped understand where the general model failed for these groups and highlighted some ecological differences in how predator groups choose their prey (Appendix S6). For example, while the match in nesting habitat and activity time of the predator and the prey were not among the important predictors

in the general model, these predictors were particularly important for some groups. Woodpeckers (*Piciformes*) tend to feed on prey that share similar nesting habitats and are active at similar times. Conversely, fowl (*Galliformes* and *Anseriformes*), pelicans (*Pelicaniformes*) and storks (*Ciconiformes*) tend to feed on prey that nest in different habitats and active at different times during the day. These are not surprising results: woodpeckers generally both forage and nest in and on trees, while many fowl and storks forage on water but nest in different habitats (Svensson & Grant 2009). We also found that the prey of other predator groups, such as carnivores (*Carnivora*), falcons (*Falconiformes*), and owls (*Strigiformes*), were harder to predict on average. These groups include many relatively generalist predators that feed on many prey types: small herbivores, but also ungulates, birds, or other carnivores (Svensson & Grant 2009; Hackländer & Zachos 2020). The functional diversity of the diet of generalist predators is larger which make it harder to find general matching rules that explain how these generalist predators choose their prey. Indeed, we found that the interactions of generalist species are harder to predict than that of specialists. Interestingly, predicting specialization has been challenging for other kinds of interactions (Blüthgen *et al.* 2008; Calatayud *et al.* 2016; Olalla-Tárraga *et al.* 2017). Although predicting trophic specialization appeared to not be an issue here for terrestrial vertebrates, further research is needed to understand the role of traits in predicting specialist and generalist interactions.

The field of species interaction and ecological network prediction has been very active in recent years (Strydom *et al.* 2021b). We presented a predictive model that aims to identify general relationships between traits and food web interactions and be used for multi-clade vertebrate groups with very few trophic interaction data. For systems in which we have more information, other approaches are possible. First, we showed that predators tend to vary in how they choose their prey between taxa, making hierarchical models good candidates to improve predictions. Hierarchical models allow some variation in the regression coefficients between groups (Ovaskainen *et al.* 2017; Gelman *et al.* 2020).

Because how species choose their prey tends to be evolutionary conserved, phylogenetic relationships could inform how regression coefficients correlate across clades (Gómez *et al.* 2010). Second, phylogenetic relationships can directly make predictions given enough interaction data (Elmasri *et al.* 2020), or to transfer species interaction knowledge between systems (Strydom *et al.* 2021a). Third, machine learning algorithms have been used to predict interactions within networks (i.e. in sample prediction) and oftentimes outperformed linear models (e.g., Desjardins-Proulx *et al.* 2017; Pichler *et al.* 2020), but often rely on a larger volume of data, and ecological inference can be less straightforward. There are also advantages to having relatively simplistic linear responses in terms of inference and the potential for better out-of-sample performance (Wenger & Olden 2012). Finally, bio-energetic models can infer energy fluxes between organisms and help quantify food webs using body mass, metabolic demands, and energy loss (Berlow *et al.* 2009; Rall *et al.* 2012). A promising avenue of research is to combine trait-based models to other methods, such as expert elicitation or bio-energetic models, to downscale regional metaweb into local quantitative food webs (Rall *et al.* 2012; Bode *et al.* 2017). This would allow us to investigate how food webs and species interactions vary in space and time, and to forecast the consequences of global changes on the composition and structure of ecosystems.

## Acknowledgements

We acknowledge that this study was conducted on unceded land which has long served as a site of meeting and exchange amongst Indigenous peoples, including the Haudenosaunee and Anishinabeg nations. We thank DC's PhD supervisory committee members Dominique Gravel, Timothée Poisot and Brian Leung, and members of the FutureWeb working group for useful comments. This research was supported by an NSERC Discovery Grant (NSERC RGPIN-2019-05771), the 2017–2018 Belmont

Forum and BiodivERsA joint call for research proposals, under the BiodivScen ERA-Net COFUND program, and with the funding organizations Agence Nationale pour la Recherche (FutureWeb, ANR-18-EBI4-0009). We declare we have no potential sources of conflict of interest.

## References

- Albouy, C., Archambault, P., Appeltans, W., Araújo, M.B., Beauchesne, D., Cazelles, K., *et al.* (2019). The marine fish food web is globally connected. *Nat Ecol Evol*, 3, 1153–1161.
- Bartomeus, I., Gravel, D., Tylianakis, J.M., Aizen, M.A., Dickie, I.A. & Bernard-Verdier, M. (2016). A common framework for identifying linkage rules across different types of interactions. *Funct. Ecol.*, 30, 1894–1903.
- Berlow, E.L., Dunne, J.A., Martinez, N.D., Stark, P.B., Williams, R.J. & Brose, U. (2009). Simple prediction of interaction strengths in complex food webs. *PNAS*, 106, 187–19
- Blüthgen, N., Fründ, J., Vázquez, D.P. & Menzel, F. (2008). What Do Interaction Network Metrics Tell Us About Specialization and Biological Traits. *Ecology*, 89, 3387–3399.
- Bode, M., Baker, C.M., Benshemesh, J., Burnard, T., Rumpff, L., Hauser, C.E., *et al.* (2017). Revealing beliefs: using ensemble ecosystem modelling to extrapolate expert beliefs to novel ecological scenarios. *Methods Ecol. Evol.*, 8, 1012–1021.
- Calatayud, J., Hórreo, J.L., Madrigal-González, J., Migeon, A., Rodríguez, M.Á., Magalhães, S., *et al.* (2016). Geography and major host evolutionary transitions shape the resource use of plant parasites. *PNAS*, 113, 9840–9845.
- Cameron, E.K., Sundqvist, M.K., Keith, S.A., CaraDonna, P.J., Mousing, E.A., Nilsson, K.A., *et al.* (2019). Uneven global distribution of food web studies under climate change. *Ecosphere*, 10, e02645.
- Canard, E., Mouquet, N., Marescot, L., Gaston, K.J., Gravel, D. & Mouillot, D. (2012). Emergence of Structural Patterns in Neutral Trophic Networks. *PLoS One*, 7, e38295.

- Chacoff, N.P., Vázquez, D.P., Lomáscolo, S.B., Stevani, E.L., Dorado, J. & Padrón, B. (2012). Evaluating sampling completeness in a desert plant–pollinator network. *J. Anim. Ecol.*, 81, 190–200.
- Deichmann, J.L., Duellman, W.E. & Williamson, G.B. (2008). Predicting Biomass from Snout–Vent Length in New World Frogs. *J. Herpetol.*, 42, 238–245.
- Desjardins-Proulx, P., Laigle, I., Poisot, T. & Gravel, D. (2017). Ecological interactions and the Netflix problem. *PeerJ*, 5, e3644.
- Dunne, J. (2006). The network structure of food webs. In: *Ecological Networks: Linking Structure to Dynamics in Food Webs*. pp. 27–86.
- Eklöf, A., Jacob, U., Kopp, J., Bosch, J., Castro-Urgal, R., Chacoff, N.P., et al. (2013). The dimensionality of ecological networks. *Ecol. Lett.*, 16, 577–583.
- Elmasri, M., Farrell, M.J., Davies, T.J. & Stephens, D.A. (2020). A hierarchical Bayesian model for predicting ecological interactions using scaled evolutionary relationships. *Ann. Appl. Stat.*, 14, 221–240.
- Estes, J.A., Terborgh, J., Brashares, J.S., Power, M.E., Berger, J., Bond, W.J., et al. (2011). Trophic Downgrading of Planet Earth. *Science*, 333, 301–306.
- Gelman, A. (2008). Scaling regression inputs by dividing by two standard deviations. *Stat Med.*, 27, 2865–2873.
- Gelman, A., Hill, J. & Vehtari, A. (2020). *Regression and Other Stories*. Cambridge University Press.
- Gelman, A. & Rubin, D.B. (1992). Inference from Iterative Simulation Using Multiple Sequences. *Stat. Sci.*, 7, 457–472.
- Golding, N. (2019). greta: simple and scalable statistical modelling in R. *JOSS*, 4, 1601.
- Gómez, J.M., Verdú, M. & Perfectti, F. (2010). Ecological interactions are evolutionarily conserved across the entire tree of life. *Nature*, 465, 918–921.

- Gravel, D., Albouy, C. & Thuiller, W. (2016). The meaning of functional trait composition of food webs for ecosystem functioning. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.*, 371, 20150268.
- Gravel, D., Poisot, T., Albouy, C., Velez, L. & Mouillot, D. (2013). Inferring food web structure from predator-prey body size relationships. *Methods Ecol. Evol.*, 4, 1083–1090.
- Hackländer, K., & Zachos, F. E. (2020). *Mammals of Europe: Past, Present, and Future*. Springer.
- Harvey, E., Gounand, I., Ward, C.L. & Altermatt, F. (2017). Bridging ecology and conservation: from ecological networks to ecosystem function. *J Appl Ecol*, 54, 371–379.
- Hortal, J., de Bello, F., Diniz-Filho, J.A.F., Lewinsohn, T.M., Lobo, J.M. & Ladle, R.J. (2015). Seven Shortfalls that Beset Large-Scale Knowledge of Biodiversity. *Annu Rev Ecol Evol Syst*, 46, 523–549.
- Jordano, P. (2016). Sampling networks of ecological interactions. *Funct. Ecol.*, 30, 1883–1893.
- Laigle, I., Aubin, I., Digel, C., Brose, U., Boulangeat, I. & Gravel, D. (2018). Species traits as drivers of food web structure. *Oikos*, 127, 316–326.
- Laliberté, E. & Tylianakis, J.M. (2010). Deforestation homogenizes tropical parasitoid–host networks. *Ecology*, 91, 1740–1747.
- Maiorano, L., Amori, G., Capula, M., Falcucci, A., Masi, M., Montemaggiori, A., et al. (2013). Threats from Climate Change to Terrestrial Vertebrate Hotspots in Europe. *PLoS One*, 8.
- Maiorano, L., Montemaggiori, A., Ficetola, G.F., O'Connor, L. & Thuiller, W. (2020). TETRA EU 1.0: A species level trophic metaweb of European tetrapods. *Glob. Ecol. Biogeogr.*, 29, 1452-1457.
- McDonald-Madden, E., Sabbadin, R., Game, E.T., Baxter, P.W.J., Chadès, I. & Possingham, H.P. (2016). Using food-web theory to conserve ecosystems. *Nat Commun*, 7, 1–8.
- Morales-Castilla, I., Matias, M.G., Gravel, D. & Araújo, M.B. (2015). Inferring biotic interactions from proxies. *Trends Ecol. Evol.*, 30, 347–356.

- O'Connor, L.M.J., Pollock, L.J., Braga, J., Ficetola, G.F., Maiorano, L., Martinez-Almoyna, C., *et al.* (2020). Unveiling the food webs of tetrapods across Europe through the prism of the Eltonian niche. *J. Biogeogr.*, 47, 181–192.
- Olalla-Tárraga, M.Á., González-Suárez, M., Bernardo-Madrid, R., Revilla, E. & Villalobos, F. (2017). Contrasting evidence of phylogenetic trophic niche conservatism in mammals worldwide. *J. Biogeogr.*, 44, 99–110.
- Ovaskainen, O., Tikhonov, G., Norberg, A., Blanchet, F.G., Duan, L., Dunson, D., *et al.* (2017). How to make more out of community data? A conceptual framework and its implementation as models and software. *Ecol. Lett.*, 20, 561–576.
- Pellissier, L., Albouy, C., Bascompte, J., Farwig, N., Graham, C., Loreau, M., *et al.* (2018). Comparing species interaction networks along environmental gradients. *Biol. Rev.*, 93, 785–800.
- Pichler, M., Boreux, V., Klein, A.-M., Schleuning, M. & Hartig, F. (2020). Machine learning algorithms to infer trait-matching and predict species interactions in ecological networks. *Methods Ecol. Evol.*, 11, 281–293.
- Piechnik, D.A., Lawler, S.P. & Martinez, N.D. (2008). Food-web assembly during a classic biogeographic study: species’ “trophic breadth” corresponds to colonization order. *Oikos*, 117, 665–674.
- Poelen, J.H., Simons, J.D. & Mungall, C.J. (2014). Global biotic interactions: An open infrastructure to share and analyze species-interaction datasets. *Ecol. Inform.*, 24, 148–159.
- Poisot, T., Bergeron, G., Cazelles, K., Dallas, T., Gravel, D., MacDonald, A., *et al.* (2021). Global knowledge gaps in species interaction networks data. *J Biogeogr.*, 48, 1552–1563.
- Poisot, T., Stouffer, D.B. & Gravel, D. (2015). Beyond species: why ecological interaction networks vary through space and time. *Oikos*, 124, 243–251.

- Pollock, L.J., O'Connor, L.M.J., Mokany, K., Rosauer, D.F., Talluto, M.V. & Thuiller, W. (2020). Protecting Biodiversity (in All Its Complexity): New Models and Methods. *Trends Ecol. Evol.*, 35, 1119–1128.
- Rall, B.C., Brose, U., Hartvig, M., Kalinkat, G., Schwarzmüller, F., Vucic-Pestic, O., et al. (2012). Universal temperature and body-mass scaling of feeding rates. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.*, 367, 2923–2934.
- Roberts, D.R., Bahn, V., Ciuti, S., Boyce, M.S., Elith, J., Guillera-Arroita, G., et al. (2017). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*, 40, 913–929.
- Rohr, R.P., Naisbit, R.E., Mazza, C. & Bersier, L.-F. (2016). Matching–centrality decomposition and the forecasting of new links in networks. *Proc. Royal Soc. B*, 283, 20152702.
- Santini, L., Benítez-López, A., Maiorano, L., Čengić, M. & Huijbregts, M.A.J. (2021). Assessing the reliability of species distribution projections in climate change research. *Divers. Distrib.*, 27, 1035–1050.
- Strydom, T., Bouskila, S., Banville, F., Barros, C., Caron, D., Farrell, M.J., et al. (2021a). Food web reconstruction through phylogenetic transfer of low-rank network representation. *EcoEvoRxiv*.
- Strydom, T., Catchen, M.D., Banville, F., Caron, D., Dansereau, G., Desjardins-Proulx, P., et al. (2021b). A roadmap towards predicting species interaction networks (across space and time). *Philos Trans R Soc Lond B Biol Sci*, 376, 20210063.
- Svensson, L. & Grant, P.J. (2009). *Birds of Europe*. Princeton field guides. 2nd ed. Princeton University Press, Princeton.
- Thompson, R.M., Brose, U., Dunne, J.A., Hall, R.O., Hladyz, S., Kitching, R.L., et al. (2012). Food webs: reconciling the structure and function of biodiversity. *Trends Ecol. Evol.*, 27, 689–697.

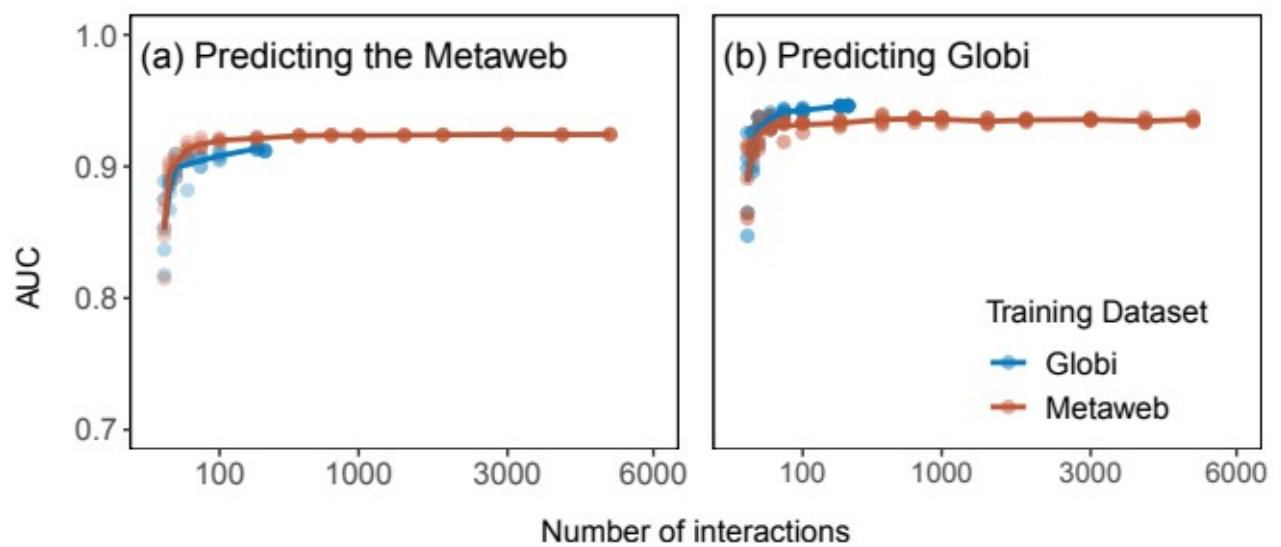
- Thuiller, W., Maiorano, L., Mazel, F., Guilhaumon, F., Ficetola, G.F., Lavergne, S., *et al.* (2015). Conserving the functional and phylogenetic trees of life of European tetrapods. *Philos. Trans. R. Soc. Lond., B, Biol. Sci*, 370, 20140005.
- Valiente-Banuet, A., Aizen, M.A., Alcántara, J.M., Arroyo, J., Cocucci, A., Galetti, M., *et al.* (2015). Beyond species loss: the extinction of ecological interactions in a changing world. *Funct. Ecol.*, 29, 299–307.
- Vesk, P.A., Morris, W.K., Neal, W.C., Mokany, K. & Pollock, L.J. (2021). Transferability of trait-based species distribution models. *Ecography*, 44, 134–147.
- Wenger, S.J. & Olden, J.D. (2012). Assessing transferability of ecological models: an underappreciated aspect of statistical validation. *Methods Ecol. Evol.*, 3, 260–267.
- Williams, R.J. & Martinez, N.D. (2000). Simple rules yield complex food webs. *Nature*, 404, 180–183.
- Wisz, M.S., Pottier, J., Kissling, W.D., Pellissier, L., Lenoir, J., Damgaard, C.F., *et al.* (2013). The role of biotic interactions in shaping distributions and realised assemblages of species: implications for species distribution modelling. *Biol. Rev.*, 88, 15–30.
- Yates, K.L., Bouchet, P.J., Caley, M.J., Mengersen, K., Randin, C.F., Parnell, S., *et al.* (2018). Outstanding Challenges in the Transferability of Ecological Models. *Trends Ecol. Evol.*, 33, 790–802.
- Zhao, Q., Brink, P.J.V. den, Carpentier, C., Wang, Y.X.G., Rodríguez-Sánchez, P., Xu, C., *et al.* (2019). Horizontal and vertical diversity jointly shape food web stability against small and large perturbations. *Ecol. Lett.*, 22, 1152–1162.

## Table and figures

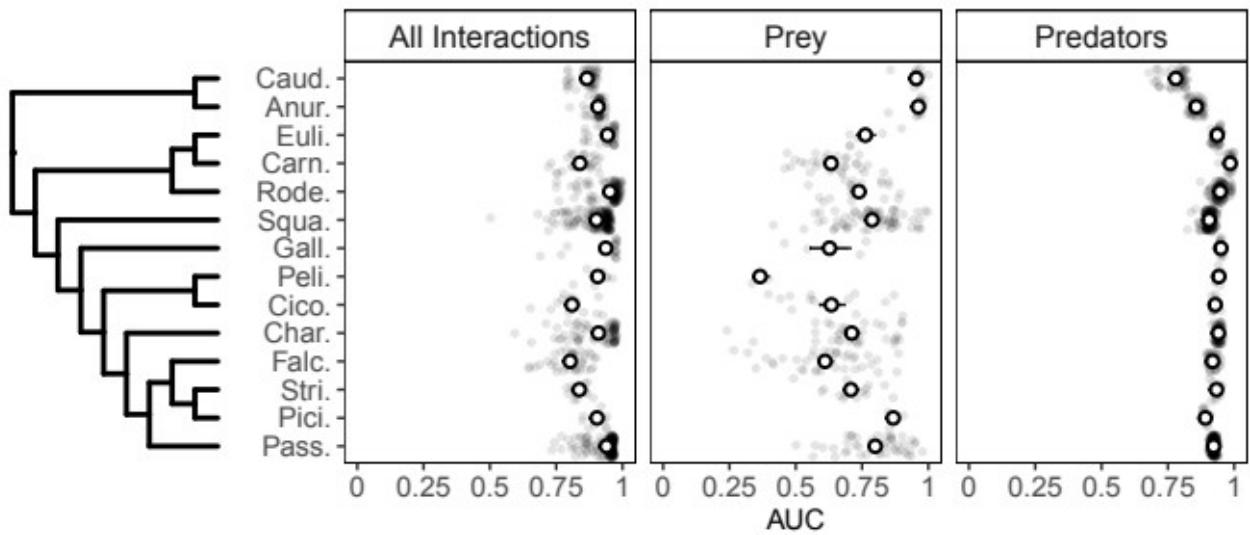
**Table 1:** Variables used to predict trophic interactions between all species pairs. The variables are grouped as foraging traits, vulnerability traits and matching traits. BM stands for body mass.

	Foraging		Vulnerability		Matching		
	Predator body mass	Predator foraging behavior	Prey body mass	Diet	Activity time	Nesting habitat	Body mass difference
Functional traits	Body Mass of the predator	Foraging behavior of the predator	Body Mass of the prey	Diet of the predator and the class of the prey	Activity time of the prey and predator	Nesting habitats of the prey and the predator	Body mass of the prey and the predator
Operation	log		log	Does the diet of the predator match the type of prey (1) or not (0) ?	Jaccard similarity*	Jaccard similarity*	$(\log(\text{BM}_{\text{predator}}) - \log(\text{BM}_{\text{prey}}))^2$
Variable type	Continuous	4 binary variables	Continuous	Binary	Continuous	Continuous	Continuous

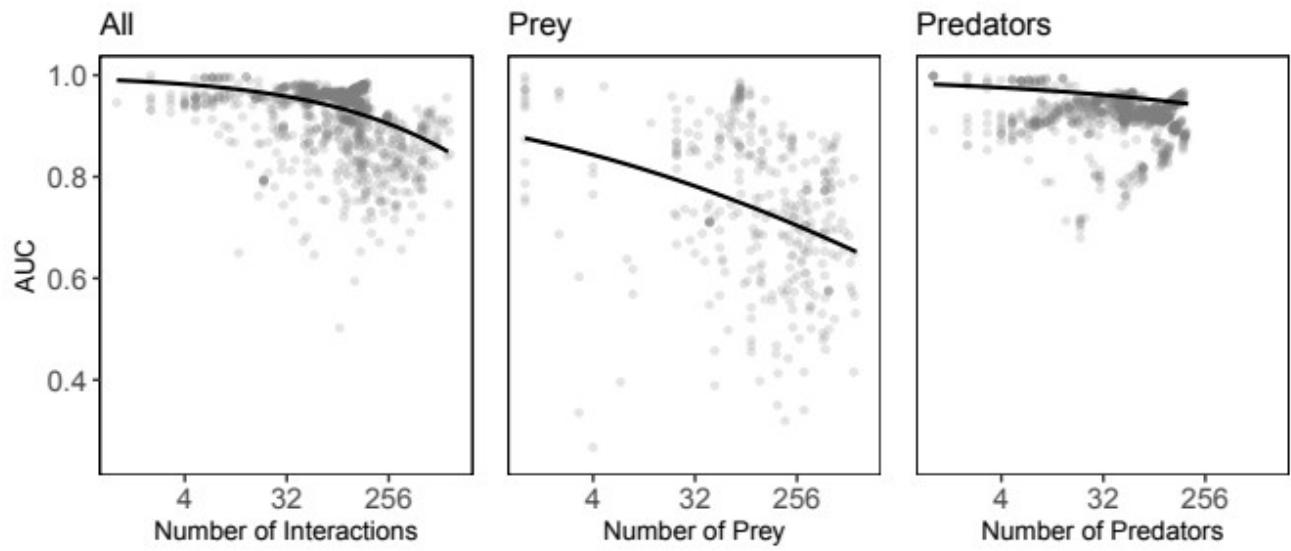
\*We calculated the Jaccard similarity coefficient for the Activity time and Nesting habitat of the predator and the prey. Activity time and Nesting habitat are a set of binary variables (Appendix S1).



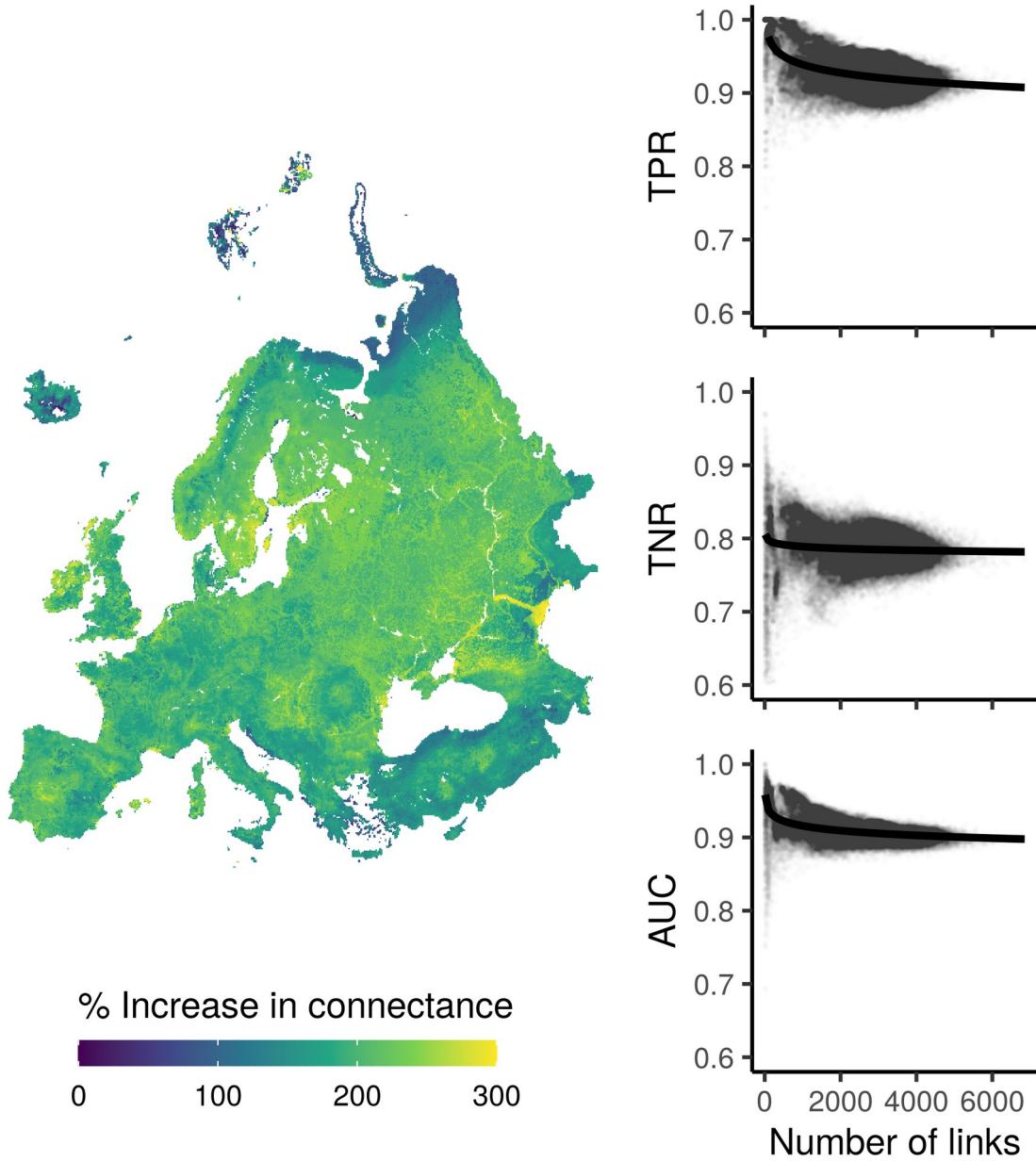
**Figure 1: Effect of the number of interactions sampled on the predictive performance of the model.** In (a), we measured performance using the European Metaweb validation dataset. In (b), we measured performance with interactions using the GloBI validation dataset. In red and blue are the predictive models trained with interactions sampled from the European metaweb and GloBI, respectively.



**Figure 2: Differences in predictive performance among groups.** Predictive performance is measured with the AUC. Each grey point is the predictive performance of the general model for a single species. The white points are the group mean, with the associated standard error. From left to right, the panels represent the performance of the general model to predict all trophic interactions, the prey, and the predators of the focal species. From top to bottom, the groups are Caudata, Anura, Eulipotyphla, Carnivora, Rodentia, Squamata, Galliformes, Pelecaniformes, Ciconiiformes, Charadriiformes, Falconiformes, Strigiformes, Piciformes, and Passeriformes.



**Figure 3: Effect of generality on model performance.** Each grey point is the predictive performance of the master model to predict all interactions (left;  $GLMM_{all}$ ), the prey (center;  $GLMM_{prey}$ ), and the predators (right;  $GLMM_{predator}$ ) of a single species. We measured generality as the number of interactions (left), the number of prey (center), and the number of predators (right) of a given species. The trend lines are the mean effect of the scaled generality (log transformed) on the logit-AUC.



**Figure 4: Spatial variation in predictability of local food webs.** The left panel shows the overestimation of connectance of each 10km-cell food web. The right panels show the true positive rate (TPR; top), the true negative rate (TNR; middle), and area under the receiver operating curve (AUC; bottom) in relation to the complexity of the local web.

## Bridging chapter 1 and chapter 2

In chapter 1, I ask how well can we predict the entire set of potential predator-prey interactions among terrestrial vertebrates when we have a subset of all interactions. This type of predictions represents interpolation as we estimate interactions within the same community from which the data used to calibrate the model comes from. I showed that even with a fraction of all interactions, we are able to estimate the full European vertebrate food web reasonably well. These findings show the potential of trait-based models to fill some of the knowledge gaps on species interactions.

However, in many ecosystems, we have little to no data on how species interact. In these instances, we need to estimate the food web using a model calibrated on data from other, better documented, ecosystems. The success of the extrapolation of trait-based models rely on how general are the relationships between the traits of predators and prey across taxa and ecosystems. This question is the focus of chapter 2 which test the transferability of trait-based model of trophic interactions across four contrasting food webs of terrestrial vertebrates: Northern Québec and Labrador, entire Europe, the eastern Pyrenees, and the Serengeti Savannah.

In chapter 1, I showed that the number of interactions tended to be overestimated, but I did not investigate other food web properties. In chapter 2, I also test whether interaction predictions scale-up to predict higher-level properties (i.e., species positions within the food web and food web properties). If higher-level properties are well predicted by interaction models, it would indicate that food webs are mainly driven by species-level constraints. In contrast, if higher-level properties are not well predicted, it would indicate that food webs are also driven by other, higher-level factors, such as dynamical, environmental, spatial, or temporal constraints.

# **Chapter 2: Trait-matching models predict pairwise interactions across regions, not food web properties**

**Dominique Caron<sup>1,2</sup>, Ulrich Brose<sup>3,4</sup>, Miguel Lurgi<sup>5,6</sup>, F., Guillaume Blanchet<sup>2,7,8,9</sup>,  
Dominique Gravel<sup>2,7</sup>, Laura J. Pollock<sup>1,2</sup>**

## **Affiliations:**

<sup>1</sup> Department of Biology, McGill University, Montreal, QC, Canada

<sup>2</sup> Quebec Centre for Biodiversity Sciences, Montreal, QC, Canada

<sup>3</sup> German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Leipzig, Germany

<sup>4</sup> Institute of Biodiversity, Friedrich Schiller University Jena, Jena, Germany

<sup>5</sup> Department of Biosciences, Swansea University, Singleton Park, UK

<sup>6</sup> Centre for Biodiversity Theory and Modelling, Theoretical and Experimental Ecology Station, CNRS, Moulis, France

<sup>7</sup> Département de biologie, Université de Sherbrooke, Sherbrooke, Quebec, Canada

<sup>8</sup> Département de mathématiques, Université de Sherbrooke, Sherbrooke, Quebec, Canada

<sup>9</sup> Département des sciences de la santé communautaire, Université de Sherbrooke, Sherbrooke, Quebec, Canada

The following has been published in *Global Ecology and Biogeography* under the reference below:

Caron, D., Brose, U., Lurgi, M., Blanchet, F. G., Gravel, D., & Pollock, L. J. (2024). Trait-matching models predict pairwise interactions across regions, not food web properties. *Global Ecology and Biogeography*, e13807. <https://doi.org/10.1111/geb.13807>

Reproduced here under the terms of the Creative Commons Attribution License.

## Abstract

**Aim:** Food webs are essential for understanding how ecosystems function. However, empirical data on the interactions that make up these ecological networks are lacking for most taxa in most ecosystems. Fortunately, they can be estimated using trait-based models to some degree. We test how well these models can extrapolate to new ecological communities (necessary for the success of these models to fill data gaps) both in terms of pairwise predator-prey interactions and higher-level food web attributes (i.e., species position, food web-level properties).

**Location:** Canada, Europe, Tanzania.

**Time period:** Current.

**Major taxa studied:** Terrestrial vertebrates.

**Methods:** We train trait-based models of pairwise trophic interactions on four independent vertebrate food webs (Canadian tundra, Serengeti, alpine south-eastern Pyrenees, and entire Europe) and evaluate how well these models predict pairwise interactions and network properties for each set of food webs.

**Results:** We find that, overall, trait-based models predict most interactions and their absence correctly, even across highly contrasting environments. Performance declined with environmental and phylogenetic distance from within food webs ( $AUC > 0.90$ ) to between webs, such as the strong contrast observed for the tundra-Serengeti ecosystems ( $AUC > 0.75$ ). Network metrics were less well-predicted than single interactions by our models with predicted food webs being more connected, less modular, and with higher mean trophic levels than observed.

**Main conclusions:** Theory predicts that the variability observed in food webs can be explained by differences in trait distributions and trait-matching relationships. Our finding that trait-based models can predict many trophic interactions, even in contrasting environments, suggests that there are general constraints on interactions and that trait-based methods can serve as a first

approximation of food webs in unknown areas. However, food webs are more than the sum of their parts, and predicting network attributes will likely require models that simultaneously predict individual interactions and community constraints.

**Key Words:** ecological predictions; food web; model transferability; terrestrial vertebrates; trait matching; trophic interactions

## Introduction

Ecosystem functions (e.g., energy flows and material cycling) and community stability depend on the trophic relationships that link species within a community (Harvey et al., 2017). Despite the importance of food webs for understanding ecosystem structure and dynamics, recognized over the last 80 years (Lindeman, 1942), we still face major challenges when developing accurate descriptions of natural food webs. One major obstacle is the lack of trophic interaction data across most locations and taxa (Poisot et al., 2021) lies at the core of the fundamental technical and practical challenges in food web ecology. The difficulty is that, to detect an interaction, one must observe individuals of both species while interacting (Jordano, 2016). This means that trophic interactions are substantially more difficult to observe than species, which is in itself a biodiversity data shortfall. Adding to the detection issues, the number of possible interactions within food webs increases quadratically with the number of species resulting in extremely large sets of potential interactions in realistic communities. Observing all possible interactions among species within a food web is very difficult even in species-poor ecosystems, so a systematic approach for predicting interactions with available data is essential.

In response to this need, models have been developed that predict interactions based on food web theory (Strydom et al., 2021). These models identify feasible interactions based on some combination of the traits of predators, the traits of prey, and the ‘match’ of traits between predator and prey (e.g., smaller predators eat smaller prey; Bartomeus et al., 2016). Such trait-based models have been used to predict food webs for freshwater streams (Pomeranz et al., 2019), marine fishes (Albouy et al., 2014), terrestrial systems comprising vertebrate (Caron et al., 2022; Fricke et al., 2022) and invertebrate species (Laigle et al., 2018), among others. Phylogenetic relationships are also informative since they can serve as proxy for trait

relationships, and because interactions and species' roles (i.e., species' positions in the food web) tend to be evolutionary conserved (Gómez et al., 2010; Stouffer et al., 2012).

While trait-based models have been shown to predict trophic interactions when fitted to data on species within a given ecosystem, a more challenging, yet largely untested, proposition is whether trait-based models fit on one ecosystem could predict the food web of another. If trait-matching rules determining interactions are general, then i) we should observe the same trait-interaction relationships across regions and ii) a model trained in one location should be transferable to other ecosystems. Also, given the strong data deficiency, many ecosystems may require extrapolation from other ecosystems. Initial results suggest that this is possible in the case of terrestrial mammals. Strydom et al. (2022) used a mammal phylogeny to map latent traits extracted from a European mammalian food web to predict its Canadian counterpart and were able to recover 90% of known trophic interactions among Canadian mammals without any prior information on the food web. This demonstrates the potential for transferability between vast regions, where trait-interaction relationships are expected to be general, with relatively similar bioclimatic conditions and somewhat related species (more than half of Canadian mammals have congeneric species in Europe). Given transferability should depend on the similarity of the environment and species composition of the respective food webs (Gravel et al., 2016), this raises the question of how well trait-based models could extrapolate to more contrasting environments.

A second, related question is whether these trait-based trophic interaction models can estimate aggregate food web properties such as network connectance, number of trophic levels and species position within the food web. These higher-level properties emerge from the joint effect

of trait-matching between interacting partners, the trait distribution of species composing the food web (Gravel et al., 2016), and are influenced by the spatial scale of the network (Galiana et al., 2018). Most studies are aimed at predicting either the properties of food webs (e.g., Williams & Martinez, 2008) or their pairwise interactions (e.g., Laigle et al., 2018; Pomeranz et al., 2019), but initial results that combine the two show that food web connectance was systematically overestimated across Europe by 2-4 times (Caron et al., 2022). While this result is not particularly surprising given the model was calibrated for predicting pairwise interactions, not network properties, it is still interesting to understand whether this is the case more generally. By doing this, we can begin to hypothesize whether there are fundamental differences in the scale of ecological processes that constrain food web properties (e.g., the number of feasible interactions) from those that drive pairwise interactions, which is commonly the focus of trophic interaction models (Strydom et al., 2021). This requires more knowledge of how well trait-matching models can be used to estimate a range of food web properties.

Here, we first ask whether predictive trait-based models from one region of the world can reliably predict predator-prey interactions in other regions. We estimate transferability (how well a model fit in one region predicts the other) between regions (Europe, the Pyrenees, Northern Québec and Labrador, and the Serengeti) using trait-based Bayesian hierarchical models. We expect better transferability between food webs that share more ecologically similar species and environment. To test this, we determine how transferability changes with geographic, environmental, trait, and phylogenetic distances. Second, we evaluate how well these models can be used to estimate a species' role in the network and food web structure. Given previous findings and the fact that the models only indirectly estimate these metrics, we expect that the models are better at predicting pairwise interactions than estimating the role of a species in a

network and food web properties. We similarly expect a decline in performance when models are trained and tested on different ecosystems.

## Methods

### *Food web data*

We extracted predator-prey interactions among terrestrial vertebrates (mammals, birds, amphibians and reptiles) from four well-resolved food webs (Table 1, Figure 1): the European food web of tetrapods (Maiorano et al., 2020), the Pyrenees vertebrate food web (Lurgi et al., 2012), the Northern Québec and Labrador food web (Berteaux et al., 2018), and the Serengeti food web (de Visser et al., 2011). The four food webs document the predator-prey interactions between all known terrestrial vertebrates in their respective geographical region. Trophic interaction is defined as a binary variable where 0 represents the absence and 1 the presence of a *potential* predator-prey interaction between two species (i.e., the predator could feed on the prey species). All four food webs were compiled from literature reviews and completed with expert knowledge. Since experts filled some of the missing interactions ubiquitous to observation-based food webs (Jordano, 2016), we expect presences and absences of interaction to be similarly uncertain. Nodes in the original Serengeti food web represent trophic groups including one or more vertebrate species. In this study, we assumed that species within a trophic group share the same predator and prey species. Thus, we can expect more false positives (i.e., non-interactions documented as interactions) in the Serengeti than in the other three food webs. Food webs are fully described in Appendix S1.

The species composition of the four food webs is different (Table 1). There are no amphibians or reptiles in Northern Québec and Labrador; the Pyrenees food web is dominated by birds (67%)

and mammals (23%), with very few reptiles (8%); the European food web has a comparable number of reptiles (21%) and mammals (25%), and almost half (46%) of the Serengeti food web are mammals. The Europe, Pyrenees, and Northern Québec and Labrador food webs all have comparable mean trophic levels (between 1.24 and 1.3) and connectance (between 0.02 and 0.05) but the Serengeti differed (mean trophic level: 1.61; connectance: 0.12). Here, the first trophic level (i.e., basal species) is defined as species not feeding on any other terrestrial vertebrates. In Europe, the Pyrenees, and Northern Québec and Labrador most species are basal species (e.g., insectivores, herbivores, piscivores), whereas many more species feed on terrestrial vertebrates (non-basal species) in the Serengeti.

#### *Trait data*

We extracted traits from the database compiled by Etard et al. (2020). This dataset combines species-level information from large freely available secondary trait databases (e.g., EltonTraits: Wilman et al. (2014), AmphiBIO: Oliveira et al. (2017)). Overall, the database includes traits for 6 990 amphibian, 11 634 bird, 5 380 mammal, and 10 612 reptile species. We extracted the available body mass (mean: 11 kg; range: 0.001–4 220 kg), longevity (mean: 5 885 days; range: 91–46 386 days), litter or clutch size (mean: 131 offspring; range: 1–20 000 offspring), habitat breadth (number of habitats a species uses, using level 2 of the IUCN Habitat Classification Scheme; mean: 10 habitats, range: 1–90 habitats), trophic level (3 levels: herbivore, omnivore, carnivore), activity time (2 categories: nocturnal, non-nocturnal) and habitat use (12 categories: forest, savanna, shrubland, grassland, wetland, rocky areas/cave/subterranean, desert, marine, marine/intertidal or coastal/supratidal, artificial, introduced vegetation) for every species considered in our study.

We standardized species names following the taxonomic backbone of the Global Biodiversity Information Facility (GBIF) using the function *name\_backbone* from the package *rgbif* (Chamberlain et al., 2022) in the R statistical language (R Core Team, 2022). We excluded species for which no taxonomic information or none of the traits were available. More than 80% of traits were available across species of all food webs (Table 1; Appendix S2). For remaining species, we imputed missing traits using the MissForest algorithm implemented in the *missForest* R package (Stekhoven & Buehlmann, 2012) for amphibians, birds, mammals and reptiles separately. MissForest uses random forests to iteratively predict missing data from the known data. Each random forest uses a different trait as response variable and the remaining traits as predictors.

#### *Phylogeny data*

We used published global phylogenies for birds (Jetz et al., 2012), amphibians (Jetz & Pyron, 2018), squamates (Tonini et al., 2016), turtles (Thomson et al., 2021), and mammals (Upham et al., 2019). All five phylogenies were built from molecular data and delivered as a posterior distribution of trees. We sampled 100 trees from the posterior of each phylogeny and calculated the mean cophenetic distance from these samples between all species of the four food webs. Following Letten & Cornwell (2015), we square root transformed cophenetic distances to better relate to ecological processes such as trophic interactions.

#### *Predictive models*

We trained a Bayesian hierarchical generalized linear model on each of the four food webs (Figure 1a) with trophic interactions following a Bernoulli distribution. Because Caron et al. (2022) found that trait-interaction relationships vary between predator groups, we added varying

slope and intercept terms for the taxonomic order of predators. For each model, we randomly drew 30% of the data for validation to keep the prevalence of trophic interaction in the validation subset equal to the prevalence of the entire food web. We used all predator-prey interactions of the remaining 70% of the data and an equal number of absences of interaction for calibration to have a more optimal balance of presences and absences for calibration (i.e., random undersampling).

$$L_{ij} \sim Bernoulli(p_{ij}) ,$$

$$\text{logit}(p_{ij}) = \alpha + \alpha_{\text{predator}[j]} + \sum_{k=1}^{13} (\beta_k \times T_k + \beta_{k,\text{predator}[j]} \times T_k) ,$$

where  $L_{ij}$  is the occurrence of interaction between predator  $j$  and prey  $i$ ,  $p_{ij}$  is the associated probability of interaction,  $\alpha$  is the fixed intercept and  $\beta_k$  the slope corresponding to the trait-based predictor  $T_k$  in the linear model. Similarly,  $\alpha_{\text{predator}[j]}$  is the random intercept and  $\beta_{k,\text{predator}[j]}$  the random slope corresponding to the trait-based predictor  $T_k$  for the order of predator  $j$ .

We used 13 trait-based predictors in the models: five predator traits (trophic level, body mass, habitat breadth, longevity, and clutch size), five prey traits (trophic level, body mass, habitat breadth, longevity, and clutch size), and three trait-match predictors (match in activity time, habitat use and body mass). Since the activity time trait is binary (nocturnal and non-nocturnal), activity time match is also a binary variable where a 1 means the predator and the prey share the same activity time trait. Habitat match is the Jaccard similarity index between the habitat used by the prey and the predator across the 12 habitat categories, which considers overlap in habitat and specialization to shared habitats. Body mass match is the squared difference between the log-transformed body mass of the prey and the predator because we expect predators to eat prey

within a given body mass interval (small enough to be handled, but large enough to be profitable). We log-transformed body mass, longevity, and clutch size, scaled each continuous predictor (after transformation) by subtracting out the mean and dividing by two times the standard deviation over the entire set of species so coefficients of continuous predictors are comparable to unscaled binary predictors (Gelman, 2008).

Fixed effects were drawn from a normal distribution with a mean of 0 and a standard deviation of

1. Random effects were drawn from normal distributions. We used a normal distribution with a mean of 0 and a standard deviation of 1 as prior for the mean and a Half-Cauchy distribution with a scale parameter of 5 as prior for the standard deviation of the random effects:

$$\begin{aligned}\alpha, \beta &\sim Normal(0,1), \\ \alpha_{predator(j)}, \beta_{k,predator(j)} &\sim Normal(\mu, \sigma), \\ \mu &\sim Normal(0,1), \\ \sigma &\sim HalfCauchy(0,5)\end{aligned}$$

Model parameters were estimated with Hamiltonian Monte Carlo (Neal, 2011). We ran four chains, each with 2000 warm-up iterations, followed by 2000 iterations for inference. We diagnosed convergence and adequacy with trace rank plots, posterior predictive checks, and we calculated the rank-normalized potential scale reduction factor on split chains for all runs (Vehtari et al., 2021; Appendix S6). We conducted the analyses using Stan (Carpenter et al., 2017) through the package *brms* in R (Bürkner, 2017).

### *Predicting species interactions*

We used each model to predict the food web on which it was trained and that of the other regions considered (Figure 1a). For each possible predator-prey pair, we extracted the mean of the

posterior as the probability of interaction, and the standard deviation of the posterior as the uncertainty around the predicted probability of interaction. We compared the predicted interaction probabilities to the validation subset of the same food web (within-food web predictions) and compared the predicted interaction probabilities to entire empirical food webs (between-food web predictions). We measured performance with the area under the receiver operating characteristic curve (AUC; Hanley & McNeil, 1982) and the area under the precision-recall-gain curve (AUPRG; Flach & Kull, 2015). AUC varies from 0 to 1 where 0.5 indicates that the model failed to rank interactions higher than absences of interactions (i.e., random predictions), and 1 indicates that the model systematically ranked interactions higher than non-interactions (i.e., perfect predictions). AUPRG is independent of the true negative rate, making it useful for highly imbalanced datasets, like food webs where there are many more absences than presences of interaction to predict (Saito & Rehmsmeier, 2015). AUPRG takes a value of 0 for random predictions and a value of 1 for perfect predictions.

We assessed the performance of models to predict each food web using three distance measures: geographic distance, environmental dissimilarity, and phylogenetic relatedness. To do so, we fitted hierarchical linear models with logit-transformed AUC as a function of distance measures, with random intercepts for the food web predicted and the one used for calibration (Appendix S7). We measured geographic distance as the great-circle distance between the polygon centroid delimiting the spatial domain of each food web. We quantified environmental dissimilarity using all 19 bioclimatic variables in WorldClim (Hijmans, 2021). We extracted the bioclimatic data for all pixels (10' resolution) falling within the spatial domain of the food webs and used the mean of each bioclimatic variable to calculate the bioclimatic centroid of each food web. We calculated environmental distance as the Euclidean distance between the food web bioclimatic centroids.

We used the mean cophenetic distance to the nearest taxon of every species in the predicted food web to measure phylogenetic relatedness. We averaged over all species in the predicted food web the cophenetic distance to the phylogenetically closest species in the food web used for calibration. This approach measures the amount of evolutionary history in the predicted food web undocumented by the food web on which the model was trained.

We also analyzed predictive performance at the species level. For each combination of predicted food web models (curves in Figure 1b), we measured how accurately the set of prey and predators of each species was predicted also using the AUC and AUPRG. We fitted a hierarchical linear model with species-specific performance as a function of how connected the focal species is and how distinct the focal species is to the species pool used to train the predictive model (Appendix S7). To do this, we used species normalized degree (number of interactions divided by the maximum possible number of interactions), the functional mean pairwise distance (Mouchet et al., 2010) and distance to nearest taxon (Tucker et al., 2017). We quantified functional mean pairwise distance using the average Gower distance (Gower, 1971) between the focal species and all species in the food web used for calibration. To calculate Gower distances, we used all traits available in Etard (2020) through the function *funct.dist* from the R package mFD (Magneville et al., 2022). We quantified distance to nearest taxon as the cophenetic distance between the focal species and the closest relative in the species in the food web used for model calibration.

#### *Predicting species' roles*

Next, we were interested in how well species' roles within the food web were predicted. The role of a species is determined by its position in the food web (Cirtwill et al., 2018), which we

quantified using five metrics related to the species' centrality within the food web (number of prey, number of predators, betweenness, closeness, eigenvector centrality), two metrics related to their trophic position (trophic levels and omnivory), two module-based metrics (within-module degree and participation coefficient; Guimerà & Amaral, 2005), and the motif role of each species (Stouffer et al., 2012). We detail each metric, their relation to species' role, and how they were calculated in Appendix S4.

We compared each species' role metric in empirical food webs to the species' role in predicted food webs. We quantified each metric of species' role on 100 draws of the posterior predictive distribution of predicted food webs. We used the mean value across posterior draws as the best point estimate for each metric, and the standard deviation as the measure of uncertainty. For each combination of model, predicted food web, and species' role metric, we fitted a linear regression between the predicted role and the empirical role and calculated the coefficient of determination ( $R^2$ ). We also explored prediction biases (e.g., systematic overestimation/underestimation) using the simple linear models' coefficients. We expect an intercept of zero for unbiased predictions of species' role. An intercept greater than zero would suggest an overall overestimation of the role metric, whereas an intercept less than zero would suggest a systematic underestimation of the role metric. We expect a slope of one for perfect predictions. A slope less than one would suggest that the role values at the lower range are overpredicted, whereas the role values at the upper range are underpredicted (i.e., more homogeneous values across species than in the empirical food web). A slope greater than one would suggest the opposite (i.e., more heterogeneous values across species than in the empirical food web).

### *Predicting food web properties*

Finally, we investigated how well the global properties of food webs were predicted. We selected a range of metrics commonly used to quantify food web structure and which have been shown to influence food web functioning and stability: connectance, mean trophic level, maximum trophic level, motif profile, food web diameter, number of clusters, and modularity (Borrelli, 2015; Vermaat et al., 2009). As for species' role, we evaluated these properties on the empirical food webs and compared them to the properties predicted using the mean of 100 draws of the posterior food web prediction. We detail each metric, their relation to food web function, and how they were calculated in Appendix S4.

#### *Boosted regression trees*

To make sure our results are general and not dependent on the model framework chosen (Bayesian hierarchical generalized linear models), we fitted boosted regression trees (BRTs) on each food web and ran the same analyses (Appendix S5). BRTs are a machine learning algorithm that allows non-linear relationships between predictors and the response variable (Friedman, 2001). To make results comparable, we used the exact same training/validation splits and the same predictor variables to train and measure performance of BRTs and Bayesian hierarchical generalized linear models.

## **Results**

#### *Predicting pairwise interactions*

For all food webs, pairwise interactions were better predicted by the model trained on the same food web (within-food web predictions) than by models trained on other food webs (between-food web predictions; Table 2). For within-food web predictions, AUC ranged from 0.92 and 0.96 and AUPRG ranged from 0.93 and 0.96. Model performance was also good (AUC > 0.82;

AUPRG > 0.84) for transfer between the Europe, Pyrenees, and Northern Québec and Labrador food webs, except for the model trained on the Pyrenees web when predicting the Europe web. For this case, AUC was high (0.89) whereas AUPRG (0.63) was comparatively small, suggesting that the model was much better at discriminating the absences than presences of interactions. Models did not transfer as well from and to the Serengeti food web, but performance was still good (AUC > 0.75; AUPRG > 0.73). The transferability of the Serengeti food web was expected to be more challenging because nodes do not represent species (as is the case for the other food webs), but trophic groups, as mentioned in the *Food web data* subsection. Overall, removing the Serengeti food web, or keeping trophic groups as the nodes of the food web did not qualitatively changed our findings (Appendix S10). The area under the precision-recall-gain curve, true positive rate, true negative rate, positive predictive value, and negative predictive value were all positively correlated with AUC and showed the same overall pattern (Appendix S7).

Model performance tended to decline for ecologically different food webs (Figure 2). Specifically, performance tended to decrease with environmental (direct effect estimate: -0.30, 95%CrI = [-0.90, 0.32]) and phylogenetic distance (direct effect estimate: -0.31, 95%CrI = [-0.69, 0.06]). Performance also tended to decrease with geographic distance (total effect estimate: -0.57, 95%CrI = [-0.78, -0.36]), but this effect largely disappeared after controlling for phylogenetic and environmental distances (direct effect estimate: -0.26, 95%CrI = [-0.87, 0.35]). Therefore, transferability likely decreases with geographic distance mainly because the environment and phylogeny diverged between the predicted food webs and those used to train the model.

We also found that species-specific model performance tended to decline when a focal species was ecologically different than the species pool used to train models (Figure 3). As expected, species with phylogenetically close relatives in the species pool used for calibration were, on average, better predicted than distant relatives (Figure 3a). Species-specific performance slightly decreased at low to intermediate phylogenetic distances and then dropped significantly at large distances. Models trained and predicted across classes (e.g., mammals to amphibians) had the lowest performance, as expected. This situation only occurs when the Northern Québec and Labrador model predicts the other three food webs as there are no amphibians or reptiles in Northern Québec and Labrador. Surprisingly, predictive performance remained qualitatively unchanged by trait distance (Figure 3b). We also found that interactions of specialist (i.e., species with few interactions) and generalist species (i.e., species with many interactions) were, on average, better predicted than interactions of species of intermediate specialization (Figure 3c).

#### *Predicting species' role*

Species' roles were slightly better predicted by within-food web predictions than by between-food web predictions (Figure 4). Interestingly, some measures of centrality (betweenness and closeness) were not well predicted, whereas others (number of prey and predators, eigenvector centrality) were relatively well predicted.

We also found important biases in the predictions of species' roles when we fitted linear regressions between species' roles in predicted food webs to their roles in the empirical food webs. The number of prey, predators, trophic level, omnivory, and the frequency of motif positions tended to be overestimated across species (intercept greater than 0). For other measures such as eigenvector centrality, betweenness, closeness, and module-based roles, species had

more similar values in the predicted than in the empirical food webs (Appendix S8, slopes less than 1).

### *Predicting food web properties*

Most food web properties were not well predicted by our interaction models (Figure 5). Connectance, mean and max trophic levels, and the frequency of most motifs were overpredicted, whereas modularity was slightly underpredicted. The frequency of the omnivory motif was particularly overpredicted, with the motif being often 50 times more prevalent in the predicted food webs than in the observed food webs. The mean and variance of relative errors were greater for between- compared to within-food web predictions. In general, predicted food webs were more connected, less modular, with fewer basal species, and increased, but less distinct (more omnivory), trophic levels.

### *Boosted regression trees*

We found the same general results with BRTs (Appendix S5). Overall, BRTs performed similarly to Bayesian linear model to predict interactions within and across food webs (AUC between 0.69 and 0.98; AUPRG between 0.73 and 0.96). Species' role and food web properties were also not well predicted by BRTs, and we found similar biases in predictions (Appendix S5).

## **Discussion**

Predictive models of trophic interactions have recently become central in filling the substantial knowledge gaps of how food webs vary across space and time. In the most comprehensive test so far, we evaluate how well these models can transfer to other ecosystems (i.e., can make predictions for an ecosystem using only data from another). Overall, we found that: (1) our

models were able to predict pairwise interactions across ecosystems reasonably well ( $AUC > 0.82$ ), (2) pairwise interactions were well predicted ( $AUC > 0.80$ ) for most novel species unless no species of the same class was present in the food web used for training, and (3) as expected, models were not as good at predicting food web properties, over-estimating some metrics (e.g. connectance, trophic level, omnivory) and under-estimating others (e.g. modularity), especially when extrapolating to new regions. Results suggest both optimism for the use of these models when data for some regions are lacking or future communities with no current analog, and the need for novel approaches that combine pairwise interactions with higher-order food web metrics.

Theory describes how the variation in food webs arises from two key processes: trait-interaction rules (e.g., body mass relationships between predators and their prey) and the distribution of traits in different systems (Gravel et al., 2016). Our ability to predict food web interactions across contrasting environments suggests that spatial food web variation is mainly driven by changes in the distribution of functional traits, and less so by the variation of trait-interaction relationships. These results are in line with previous research finding generalities made on trait-matching relationships across European bioregions (Caron et al., 2022), predator-prey body-size ratios within habitat, predator, and prey types (Brose et al., 2006), and the trait-interaction relationships in soil invertebrates across three forest areas in Germany (Laigle et al., 2018). The generalities we found in trait-matching relationships within major taxonomic groups (order and class) also add to previous research indicating evolutionary conservation of species interactions (Gómez et al., 2010) and role (Stouffer et al., 2012). These generalities in the trait-interaction relationships across space and within taxonomic groups demonstrate the potential for using data

from well-studied areas to predict interactions in areas we know very little about or forecast (and hindcast) food webs given new trait distributions.

As expected, pairwise trophic interaction models predict interactions, not necessarily food web properties, and here this is shown for a much wider set of metrics than previous comparisons (e.g., Caron et al., 2022). The explanation could be biological, methodological, or both. Food webs are more than a collection of independent pairwise interactions (Strydom et al., 2021) and could be constrained by ecological processes such as dispersal limitation or even area (Galiana et al., 2018), which are not included in our models. How these constraints lead to the spatial and temporal variation of ecological networks is gaining interest recently (Baiser et al., 2019; Gravel et al., 2019). While a process-based understanding of ecological constraints is beyond the scope of this study, our findings can offer some hints as to what might be causing the differences between predicted and observed network properties.

In particular, our study adds to the previous findings that connectance is over-estimated with these models, often with the number of interactions two-fold greater than observed. Caron et al. (2022) previously found similar results for the European web and showed how the over-estimation increased with species richness. Our study additionally shows that the over-estimation is even more exaggerated with extrapolation (predicting to novel conditions), in which the models are even less likely to contain the relevant ecological constraints. For example, if there is a different structure of specialists and generalists in the training and testing ecosystem, then we would expect differences in network properties. The Serengeti contained more generalist and less basal species than others, so models trained there are more likely to overestimate connectance in other regions. Our results also show that predator-prey interactions are predicted less accurately

in species with intermediate levels of specialization. This suggests that trophic niches of species that are neither specialists nor generalists are harder to define by traits and would ask for more attention when sampling interactions. Finally, our results show that predicted food web have more, but less distinct (more omnivory), trophic levels. Again, adding better species trophic level information, or giving special attention to omnivorous interactions during sampling, would help refine predicted food webs.

The overall overestimation of the number of prey and predators (Appendix S8), probably propagated through the food webs, explaining why the centrality of species was more evenly distributed, and why the predicted food webs were more connected, less modular, and with higher trophic levels. Statistically speaking, even if most interactions and absences of interactions are well predicted (high true positive and negative rates), there are often an order of magnitude more absences of interactions to predict than presences in real food webs (low connectance). Another explanation recently proposed is constraints on food webs due to the area of the food web (Galiana et al., 2018). This could influence our findings as we compared food webs with very different areas (ranging from 36 thousand km<sup>2</sup> for Serengeti to 11 million km<sup>2</sup> for Europe). However, in this case, area does not seem to be a major constraint given within-food web predictions were also biased. Additional research is needed to understand how realized food webs differ from feasible trophic interactions.

There are also many methodological possibilities for correcting biases in higher-level properties that arise from aggregating sets of independent models. This problem has been addressed extensively in the field of species distribution modelling, where individual models predict species well, but not necessarily species richness in communities (Zurell et al., 2020). Methods

have been developed to harness biases in higher-level properties to correct distribution predictions. For example, (Leung et al., (2019) combined predicted species distributions and expected species richness to quantify species and spatial biases. They correlated biases to species traits and environmental co-variates to improve predicted species distribution. Similar methods integrating predictions of interactions and networks have the potential to provide better food web predictions (Isaac et al., 2020).

Until then, these food webs remain very useful first-order approximations of food webs when there is little to no data in these ecosystems. Most food webs, even expert-based, are missing some interactions (e.g., cryptic and opportunistic interactions). And one could argue that over-estimating connections is better than under-estimating given these are candidate interactions, which can then be refined through geographic subsets, expert opinion or additional data.

Future studies could address other limitations of our study, such as the reliance on terrestrial vertebrates. The exclusion of non-vertebrates (e.g., plants, invertebrates, parasites) meant the first trophic levels were vertebrates, not primary producers, but the extension of trait-matching models to also include invertebrates can be achieved with a coarser resolution (Li et al., 2023) or with additional traits (Laigle et al., 2018). Second, the food webs we used were binary food webs. Trait-matching models predict the probabilities that a species could eat another species given they are encountering each other. Additional data, such as co-occurrence and abundance data, are needed to make predictions of realized and quantitative interactions. Third, due to the scarcity of food web data, we only had four food webs to work with. This means we only had four sets of within-food web predictions and 12 sets of between-food web predictions. This explains the large uncertainty for some of our results (e.g., Figure 2). Finally, our trait-based

predictions rely on species-level traits, ignoring intraspecific variation within and across regions. Different environmental conditions and co-evolutionary dynamics can lead to shifts in trait-interaction relationships if the shifts in the traits of prey and predators are asymmetric (Gomulkiewicz et al., 2000). It is reasonable to believe that intraspecific variation is less important in regional and potential food webs than local and realized food webs, but future studies would need to evaluate the extent and scale to which co-evolutionary dynamics influence trait-based predictions of trophic interactions and food webs.

Overall, we found that our trait-based interaction models can transfer knowledge relatively well given enough phylogenetic and environmental similarities between systems. They are useful for producing initial expectations even in contrasting ecosystems and for future communities. This, and other recent research (Brose et al., 2006; Caron et al., 2022; Laigle et al., 2018), suggests there are fundamental trait-based constraints on trophic interactions that are generalizable to some extent. Food web variation thus results from species traits rather than the ‘match’ between a predator and its prey, which has interesting ties to the broader question of how traits vary between regions and along gradients (Gravel et al., 2016). Future research could better link subfields within trait-based ecology for a comprehensive understanding of how species traits and network structure relate to their environment and enable trait-matching approach to inform restoration and re-wilding initiatives on the choice of appropriate species to re-establish former links and potentially ecosystem functions related to these interactions.

## References

- Albouy, C., Velez, L., Coll, M., Colloca, F., Loc'h, F. L., Mouillot, D., & Gravel, D. (2014). From projected species distribution to food-web structure under climate change. *Global Change Biology*, 20(3), 730–741. <https://doi.org/10.1111/gcb.12467>
- Baiser, B., Gravel, D., Cirtwill, A. R., Dunne, J. A., Fahimipour, A. K., Gilarranz, L. J., Grochow, J. A., Li, D., Martinez, N. D., McGrew, A., Poisot, T., Romanuk, T. N., Stouffer, D. B., Trotta, L. B., Valdovinos, F. S., Williams, R. J., Wood, S. A., & Yeakel, J. D. (2019). Ecogeographical rules and the macroecology of food webs. *Global Ecology and Biogeography*, 28(9), 1204–1218. <https://doi.org/10.1111/geb.12925>
- Bartomeus, I., Gravel, D., Tylianakis, J. M., Aizen, M. A., Dickie, I. A., & Bernard-Verdier, M. (2016). A common framework for identifying linkage rules across different types of interactions. *Functional Ecology*, 30(12), 1894–1903. <https://doi.org/10.1111/1365-2435.12666>
- Berteaux, D., Ropars, P., & Casajus, N. (2018). Toundra Nunavik: Matrice des relations trophiques entre espèces du Labrador et du Québec nordique, v. 1.0 (1980-2010). *Nordicana D36*. <https://doi.org/10.5885/45555CE-DA1FF11FA4254703>
- Blanchet, F. G., Cazelles, K., & Gravel, D. (2020). Co-occurrence is not evidence of ecological interactions. *Ecology Letters*, 23(7), 1050–1063. <https://doi.org/10.1111/ele.13525>
- Borrelli, J. J. (2015). Selection against instability: Stable subgraphs are most frequent in empirical food webs. *Oikos*, 124(12), 1583–1588. <https://doi.org/10.1111/oik.02176>
- Brose, U., Jonsson, T., Berlow, E. L., Warren, P., Banasek-Richter, C., Bersier, L.-F., Blanchard, J. L., Brey, T., Carpenter, S. R., Blandenier, M.-F. C., Cushing, L., Dawah, H. A., Dell, T., Edwards, F., Harper-Smith, S., Jacob, U., Ledger, M. E., Martinez, N. D., Memmott, J., ... Cohen, J. E. (2006). Consumer–Resource Body-Size Relationships in Natural Food

Webs. *Ecology*, 87(10), 2411–2417. [https://doi.org/10.1890/0012-9658\(2006\)87\[2411:CBRINF\]2.0.CO;2](https://doi.org/10.1890/0012-9658(2006)87[2411:CBRINF]2.0.CO;2)

Bürkner, P.-C. (2017). **brms**: An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software*, 80(1). <https://doi.org/10.18637/jss.v080.i01>

Canard, E., Mouquet, N., Marescot, L., Gaston, K. J., Gravel, D., & Mouillot, D. (2012).

Emergence of Structural Patterns in Neutral Trophic Networks. *PLOS ONE*, 7(8), e38295. <https://doi.org/10.1371/journal.pone.0038295>

Caron, D., Maiorano, L., Thuiller, W., & Pollock, L. J. (2022). Addressing the Eltonian shortfall with trait-based interaction models. *Ecology Letters*, ele.13966.

<https://doi.org/10.1111/ele.13966>

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., & Riddell, A. (2017). Stan: A Probabilistic Programming Language. *Journal of Statistical Software*, 76(1). <https://doi.org/10.18637/jss.v076.i01>

Chamberlain, S., Barve, V., Mcglinn, D., Oldoni, D., Desmet, P., Geffert, L., & Ram, K. (2022). *rgbif: Interface to the global biodiversity information facility API [Manual]*.

<https://CRAN.R-project.org/package=rgbif>

Cirtwill, A. R., Dalla Riva, G. V., Gaiarsa, M. P., Bimler, M. D., Cagua, E. F., Coux, C., & Dehling, D. M. (2018). A review of species role concepts in food webs. *Food Webs*, 16, e00093. <https://doi.org/10.1016/j.fooweb.2018.e00093>

Davis, J., & Goadrich, M. (2006). The relationship between Precision-Recall and ROC curves. *Proceedings of the 23rd International Conference on Machine Learning - ICML '06*, 233–240. <https://doi.org/10.1145/1143844.1143874>

de Visser, S. N., Freymann, B. P., & Olff, H. (2011). The Serengeti food web: Empirical quantification and analysis of topological changes under increasing human impact.

*Journal of Animal Ecology*, 80(2), 484–494. <https://doi.org/10.1111/j.1365-2656.2010.01787.x>

Duffy, J. E., Cardinale, B. J., France, K. E., McIntyre, P. B., Thébaud, E., & Loreau, M. (2007).

The functional role of biodiversity in ecosystems: Incorporating trophic complexity.

*Ecology Letters*, 10(6), 522–538. <https://doi.org/10.1111/j.1461-0248.2007.01037.x>

Estes, J. A., Terborgh, J., Brashares, J. S., Power, M. E., Berger, J., Bond, W. J., Carpenter, S. R.,

Essington, T. E., Holt, R. D., Jackson, J. B. C., Marquis, R. J., Oksanen, L., Oksanen, T.,

Paine, R. T., Pikitch, E. K., Ripple, W. J., Sandin, S. A., Scheffer, M., Schoener, T. W., ...

Wardle, D. A. (2011). Trophic Downgrading of Planet Earth. *Science*, 333(6040), 301–

306. <https://doi.org/10.1126/science.1205106>

Etard, A., Morrill, S., & Newbold, T. (2020). Global gaps in trait data for terrestrial vertebrates.

*Global Ecology and Biogeography*, 29(12), 2143–2158.

<https://doi.org/10.1111/geb.13184>

Flach, P., & Kull, M. (2015). Precision-Recall-Gain Curves: PR Analysis Done Right. In C.

Cortes, N. Lawrence, D. Lee, M. Sugiyama, & R. Garnett (Eds.), *Advances in Neural Information Processing Systems* (Vol. 28). Curran Associates, Inc.

[https://proceedings.neurips.cc/paper\\_files/paper/2015/file/33e8075e9970de0cfea955af4644bb2-Paper.pdf](https://proceedings.neurips.cc/paper_files/paper/2015/file/33e8075e9970de0cfea955af4644bb2-Paper.pdf)

Fricke, E. C., Hsieh, C., Middleton, O., Gorczynski, D., Cappello, C. D., Sanisidro, O., Rowan,

J., Svenning, J.-C., & Beaudrot, L. (2022). Collapse of terrestrial mammal food webs

since the Late Pleistocene. *Science*, 377(6609), 1008–1011.

<https://doi.org/10.1126/science.abn4012>

Friedman, J. H. (2001). Greedy Function Approximation: A Gradient Boosting Machine. *The*

*Annals of Statistics*, 29(5), 1189–1232. JSTOR.

- Galiana, N., Lurgi, M., Claramunt-López, B., Fortin, M.-J., Leroux, S., Cazelles, K., Gravel, D., & Montoya, J. M. (2018). The spatial scaling of species interaction networks. *Nature Ecology & Evolution*, 2(5), Article 5. <https://doi.org/10.1038/s41559-018-0517-3>
- Gelman, A. (2008). Scaling regression inputs by dividing by two standard deviations. *Statistics in Medicine*, 27(15), 2865–2873. <https://doi.org/10.1002/sim.3107>
- Gómez, J. M., Verdú, M., & Perfectti, F. (2010). Ecological interactions are evolutionarily conserved across the entire tree of life. *Nature*, 465(7300), Article 7300. <https://doi.org/10.1038/nature09113>
- Gower, J. C. (1971). A General Coefficient of Similarity and Some of Its Properties. *Biometrics*, 27(4), 857–871. <https://doi.org/10.2307/2528823>
- Gravel, D., Albouy, C., & Thuiller, W. (2016). The meaning of functional trait composition of food webs for ecosystem functioning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371(1694), 20150268. <https://doi.org/10.1098/rstb.2015.0268>
- Gravel, D., Baiser, B., Dunne, J. A., Kopelke, J.-P., Martinez, N. D., Nyman, T., Poisot, T., Stouffer, D. B., Tylianakis, J. M., Wood, S. A., & Roslin, T. (2019). Bringing Elton and Grinnell together: A quantitative framework to represent the biogeography of ecological interaction networks. *Ecography*, 42(3), 401–415. <https://doi.org/10.1111/ecog.04006>
- Guimerà, R., & Amaral, L. A. N. (2005). Cartography of complex networks: Modules and universal roles. *Journal of Statistical Mechanics: Theory and Experiment*, 2005(02), P02001. <https://doi.org/10.1088/1742-5468/2005/02/P02001>
- Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, 143(1), 29–36. <https://doi.org/10.1148/radiology.143.1.7063747>

- Harvey, E., Gounand, I., Ward, C. L., & Altermatt, F. (2017). Bridging ecology and conservation: From ecological networks to ecosystem function. *Journal of Applied Ecology*, 54(2), 371–379. <https://doi.org/10.1111/1365-2664.12769>
- Hijmans, R. J. (2021). *raster: Geographic data analysis and modeling* [Manual]. <https://CRAN.R-project.org/package=raster>
- Jetz, W., & Pyron, R. A. (2018). The interplay of past diversification and evolutionary isolation with present imperilment across the amphibian tree of life. *Nature Ecology & Evolution*, 2(5), Article 5. <https://doi.org/10.1038/s41559-018-0515-5>
- Jetz, W., Thomas, G. H., Joy, J. B., Hartmann, K., & Mooers, A. O. (2012). The global diversity of birds in space and time. *Nature*, 491(7424), Article 7424. <https://doi.org/10.1038/nature11631>
- Jordano, P. (2016). Sampling networks of ecological interactions. *Functional Ecology*, 30(12), 1883–1893. <https://doi.org/10.1111/1365-2435.12763>
- Laigle, I., Aubin, I., Digel, C., Brose, U., Boulangeat, I., & Gravel, D. (2018). Species traits as drivers of food web structure. *Oikos*, 127(2), 316–326. <https://doi.org/10.1111/oik.04712>
- Letten, A. D., & Cornwell, W. K. (2015). Trees, branches and (square) roots: Why evolutionary relatedness is not linearly related to functional distance. *Methods in Ecology and Evolution*, 6(4), 439–444. <https://doi.org/10.1111/2041-210X.12237>
- Leung, B., Hudgins, E. J., Potapova, A., & Ruiz-Jaen, M. C. (2019). A new baseline for countrywide  $\alpha$ -diversity and species distributions: Illustration using >6,000 plant species in Panama. *Ecological Applications*, 29(3), e01866. <https://doi.org/10.1002/eap.1866>
- Li, J., Luo, M., Wang, S., Gauzens, B., Hirt, M. R., Rosenbaum, B., & Brose, U. (2023). A size-constrained feeding-niche model distinguishes predation patterns between aquatic and terrestrial food webs. *Ecology Letters*, 26(1), 76–86. <https://doi.org/10.1111/ele.14134>

Lindeman, R. L. (1942). The Trophic-Dynamic Aspect of Ecology. *Ecology*, 23(4), 399–417.

<https://doi.org/10.2307/1930126>

Lurgi, M., Galiana, N., López, B. C., Joppa, L. N., & Montoya, J. M. (2014). Network complexity and species traits mediate the effects of biological invasions on dynamic food webs. *Frontiers in Ecology and Evolution*, 2.

<https://www.frontiersin.org/articles/10.3389/fevo.2014.00036>

Lurgi, M., López, B. C., & Montoya, J. M. (2012). Climate change impacts on body size and food web structure on mountain ecosystems. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1605), 3050–3057.

<https://doi.org/10.1098/rstb.2012.0239>

Magneville, C., Loiseau, N., Albouy, C., Casajus, N., Claverie, T., Escalas, A., Leprieur, F., Maire, E., Mouillot, D., & Villéger, S. (2022). mFD: An R package to compute and illustrate the multiple facets of functional diversity. *Ecography*, 2022(1).

<https://doi.org/10.1111/ecog.05904>

Maiorano, L., Montemaggiori, A., Ficetola, G. F., O'Connor, L., & Thuiller, W. (2020). TETRA-EU 1.0: A species-level trophic metaweb of European tetrapods. *Global Ecology and Biogeography*, geb.13138. <https://doi.org/10.1111/geb.13138>

Mouchet, M. A., Villéger, S., Mason, N. W. H., & Mouillot, D. (2010). Functional diversity measures: An overview of their redundancy and their ability to discriminate community assembly rules. *Functional Ecology*, 24(4), 867–876. <https://doi.org/10.1111/j.1365-2435.2010.01695.x>

Neal, R. M. (2011). MCMC Using Hamiltonian Dynamics. In *Handbook of Markov Chain Monte Carlo* (pp. 113–162). CRC Press. <https://doi.org/10.1201/b10905-7>

- Oliveira, B. F., São-Pedro, V. A., Santos-Barrera, G., Penone, C., & Costa, G. C. (2017). AmphiBIO, a global database for amphibian ecological traits. *Scientific Data*, 4(1), Article 1. <https://doi.org/10.1038/sdata.2017.123>
- Poisot, T., Bergeron, G., Cazelles, K., Dallas, T., Gravel, D., MacDonald, A., Mercier, B., Violet, C., Vissault, S., & Chapman, D. (2021). Global knowledge gaps in species interaction networks data. *Journal of Biogeography*, jbi.14127. <https://doi.org/10.1111/jbi.14127>
- Pomeranz, J. P. F., Thompson, R. M., Poisot, T., & Harding, J. S. (2019). Inferring predator–prey interactions in food webs. *Methods in Ecology and Evolution*, 10(3), 356–367. <https://doi.org/10.1111/2041-210X.13125>
- R Core Team. (2022). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Saito, T., & Rehmsmeier, M. (2015). The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets. *PLOS ONE*, 10(3), e0118432. <https://doi.org/10.1371/journal.pone.0118432>
- Stekhoven, D. J., & Buehlmann, P. (2012). MissForest—Non-parametric missing value imputation for mixed-type data. *Bioinformatics (Oxford, England)*, 28(1), 112–118.
- Stouffer, D. B., Sales-Pardo, M., Sirer, M. I., & Bascompte, J. (2012). Evolutionary Conservation of Species' Roles in Food Webs. *Science*, 335(6075), 1489–1492. <https://doi.org/10.1126/science.1216556>
- Strydom, T., Bouskila, S., Banville, F., Barros, C., Caron, D., Farrell, M. J., Fortin, M.-J., Hemming, V., Mercier, B., Pollock, L. J., Runghen, R., Dalla Riva, G. V., & Poisot, T. (2022). Food web reconstruction through phylogenetic transfer of low-rank network representation. *Methods in Ecology and Evolution*, 13(12), 2838–2849. <https://doi.org/10.1111/2041-210X.13835>

- Strydom, T., Catchen, M. D., Banville, F., Caron, D., Dansereau, G., Desjardins-Proulx, P., Forero-Muñoz, N. R., Higino, G., Mercier, B., Gonzalez, A., Gravel, D., Pollock, L., & Poisot, T. (2021). A roadmap towards predicting species interaction networks (across space and time). *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 376(1837), 20210063. <https://doi.org/10.1098/rstb.2021.0063>
- Thomson, R. C., Spinks, P. Q., & Shaffer, H. B. (2021). A global phylogeny of turtles reveals a burst of climate-associated diversification on continental margins. *Proceedings of the National Academy of Sciences*, 118(7), e2012215118.  
<https://doi.org/10.1073/pnas.2012215118>
- Tonini, J. F. R., Beard, K. H., Ferreira, R. B., Jetz, W., & Pyron, R. A. (2016). Fully-sampled phylogenies of squamates reveal evolutionary patterns in threat status. *Biological Conservation*, 204, 23–31. <https://doi.org/10.1016/j.biocon.2016.03.039>
- Tucker, C. M., Cadotte, M. W., Carvalho, S. B., Davies, T. J., Ferrier, S., Fritz, S. A., Grenyer, R., Helmus, M. R., Jin, L. S., Mooers, A. O., Pavoine, S., Purschke, O., Redding, D. W., Rosauer, D. F., Winter, M., & Mazel, F. (2017). A guide to phylogenetic metrics for conservation, community ecology and macroecology. *Biological Reviews*, 92(2), 698–715. <https://doi.org/10.1111/brv.12252>
- Upham, N. S., Esselstyn, J. A., & Jetz, W. (2019). Inferring the mammal tree: Species-level sets of phylogenies for questions in ecology, evolution, and conservation. *PLOS Biology*, 17(12), e3000494. <https://doi.org/10.1371/journal.pbio.3000494>
- Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., & Bürkner, P.-C. (2021). Rank-Normalization, Folding, and Localization: An Improved  $R^*$  for Assessing Convergence of MCMC (with Discussion). *Bayesian Analysis*, 16(2), 667–718.  
<https://doi.org/10.1214/20-BA1221>

- Vermaat, J. E., Dunne, J. A., & Gilbert, A. J. (2009). Major dimensions in food-web structure properties. *Ecology*, 90(1), 278–282. <https://doi.org/10.1890/07-0978.1>
- Wang, S., & Brose, U. (2018). Biodiversity and ecosystem functioning in food webs: The vertical diversity hypothesis. *Ecology Letters*, 21(1), 9–20.  
<https://doi.org/10.1111/ele.12865>
- Williams, R. J., & Martinez, N. D. (2008). Success and Its Limits among Structural Models of Complex Food Webs. *Journal of Animal Ecology*, 77(3), 512–519.
- Wilman, H., Belmaker, J., Simpson, J., Rosa, C. de la, Rivadeneira, M. M., & Jetz, W. (2014). EltonTraits 1.0: Species-level foraging attributes of the world's birds and mammals. *Ecology*, 95(7), 2027–2027. <https://doi.org/10.1890/13-1917.1>
- Zurell, D., Zimmermann, N. E., Gross, H., Baltensweiler, A., Sattler, T., & Wüest, R. O. (2020). Testing species assemblage predictions from stacked and joint species distribution models. *Journal of Biogeography*, 47(1), 101–113. <https://doi.org/10.1111/jbi.13608>

## **Data Accessibility Statement**

All data sources have been previously published. All data used in the analyses and relevant code are archived on the Open Science Framework repository [OSF | Interaction model Transferability](#).

## **Acknowledgements**

This research was supported by an NSERC Discovery Grant (NSERC RGPIN-2019-05771). We thank Brian Leung and Luigi Maiorano for comments on the study design and results and Timothée Poisot for discussions on model calibration and validation.

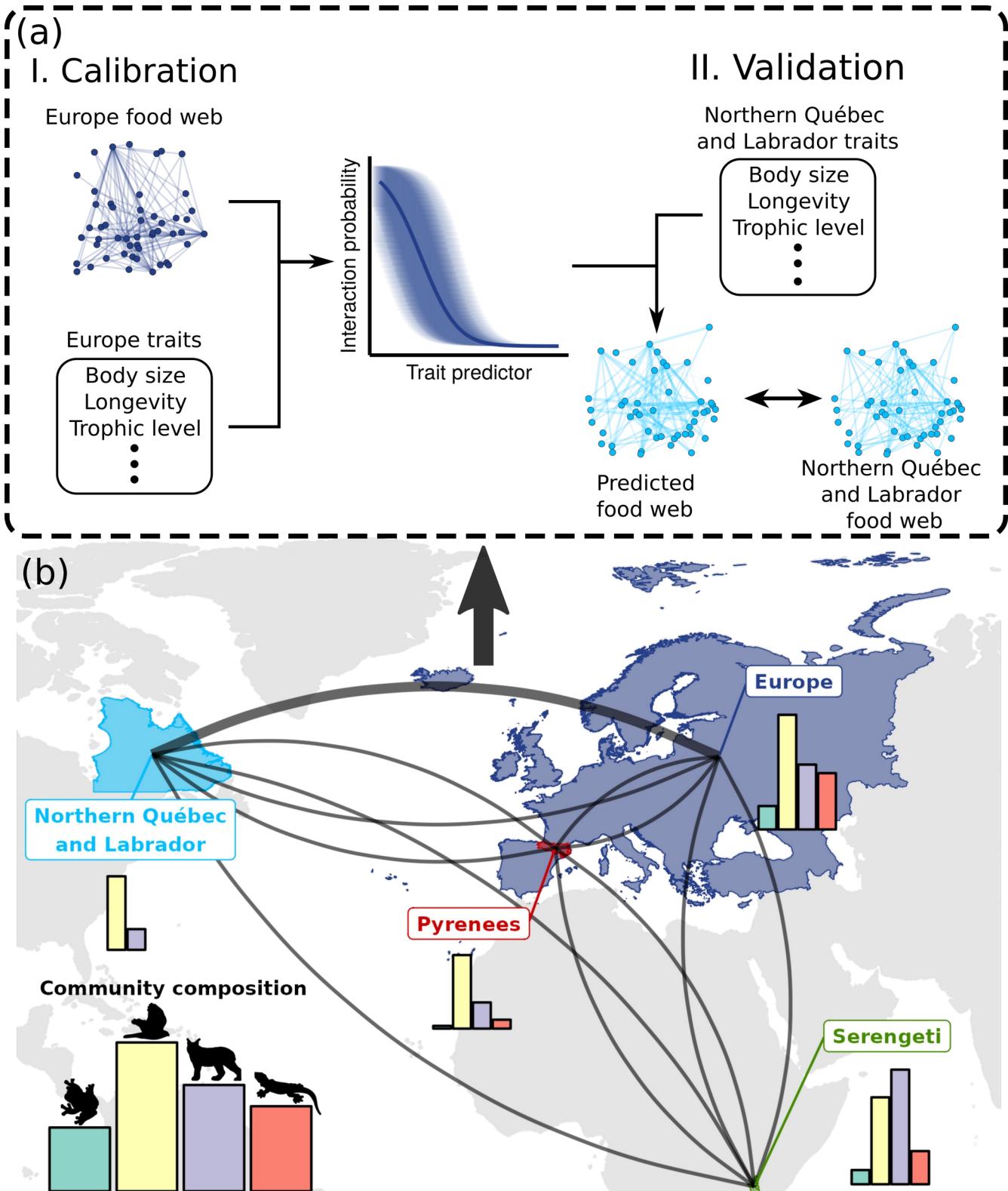
## Tables and figures

*Table 2: Summary statistics of the food webs used in this study after excluding species for which no match was found in the GBIF Backbone Taxonomy or was not documented in the trait database. Trait coverage is the percentage of traits documented in Etard et al. (2020) before imputing missing traits.*

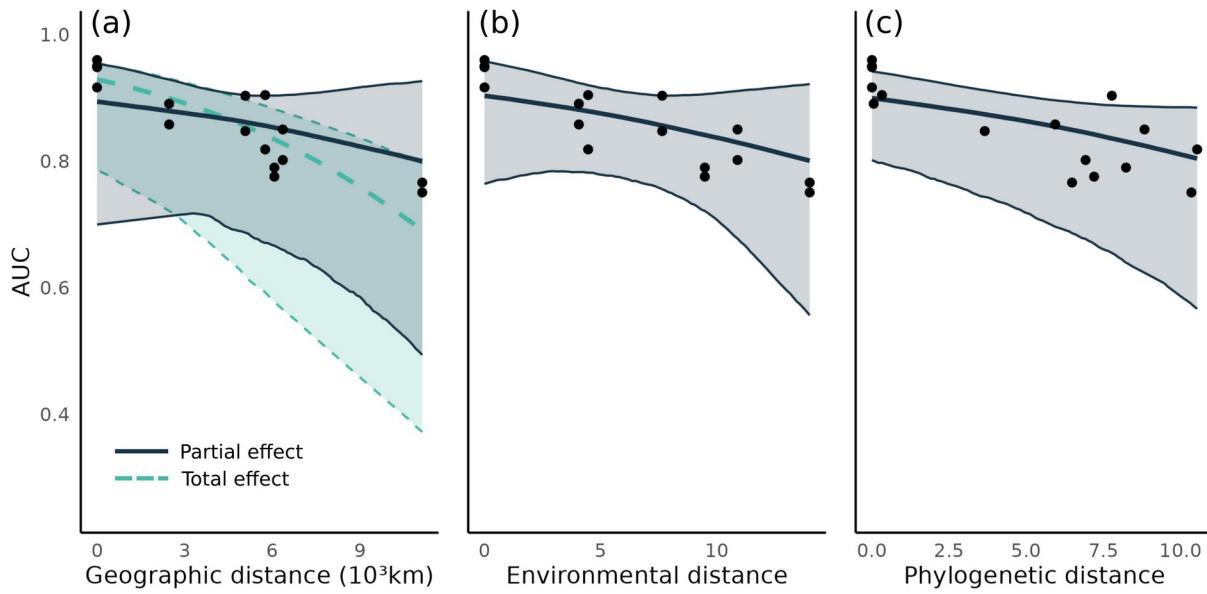
	Europe	Pyrenees	North Québec and Labrador	Serengeti
No. of species	1135	196	200	298
Prop of amphibians	0.09	0.02	0	0.06
Prop of birds	0.45	0.67	0.78	0.35
Prop of mammals	0.25	0.23	0.22	0.46
Prop of reptiles	0.21	0.08	0	0.13
No. of interactions	57 746	831	1 098	11 038
Connectance	0.05	0.02	0.03	0.12
Mean trophic level	1.24	1.25	1.3	1.61
Trait coverage (%)	83.9	92.1	96.3	81.9

*Table 3: Area under the receiver operating curve (AUC) and area under the precision-recall-gain curve (AUPRG) where each food web model is used to predict food web data.*

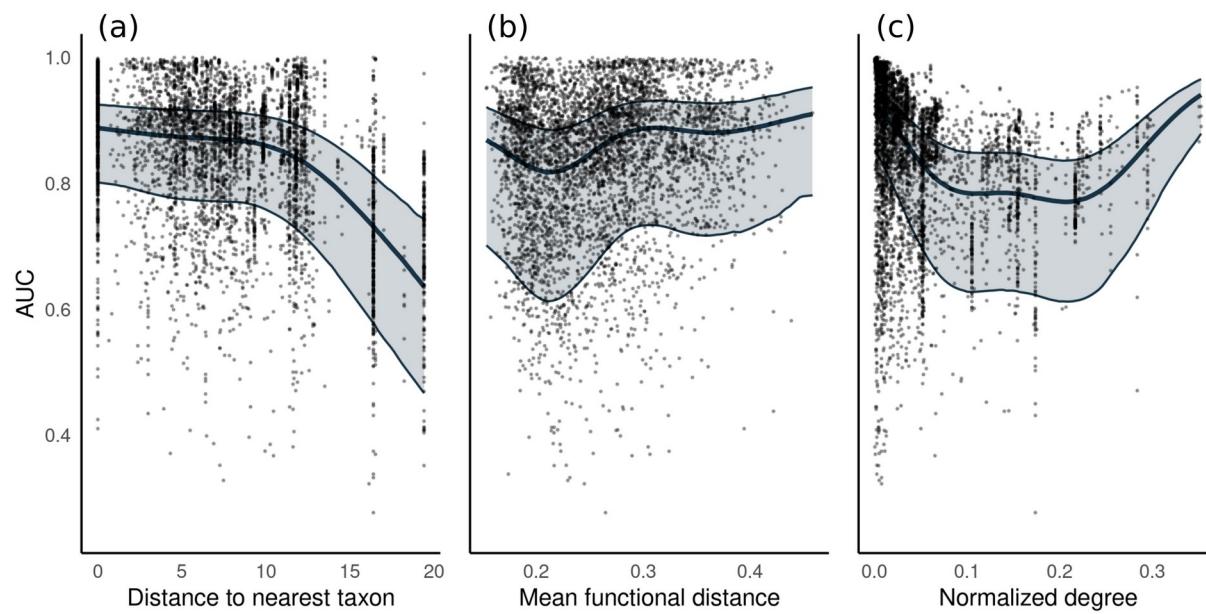
Model	Food web			
	Europe	Pyrenees	North Québec and Labrador	Serengeti
<b>AUC</b>				
Europe	<b>0.96</b>	0.89	0.9	0.8
Pyrenees	0.87	<b>0.95</b>	0.86	0.76
North Québec And Labrador	0.81	0.91	<b>0.95</b>	0.76
Serengeti	0.84	0.87	0.78	<b>0.95</b>
<b>AUPRG</b>				
Europe	<b>0.97</b>	0.66	0.86	0.84
Pyrenees	0.9	<b>0.98</b>	0.95	0.84
North Québec And Labrador	0.88	0.91	<b>0.96</b>	0.8
Serengeti	0.84	0.86	0.83	<b>0.97</b>



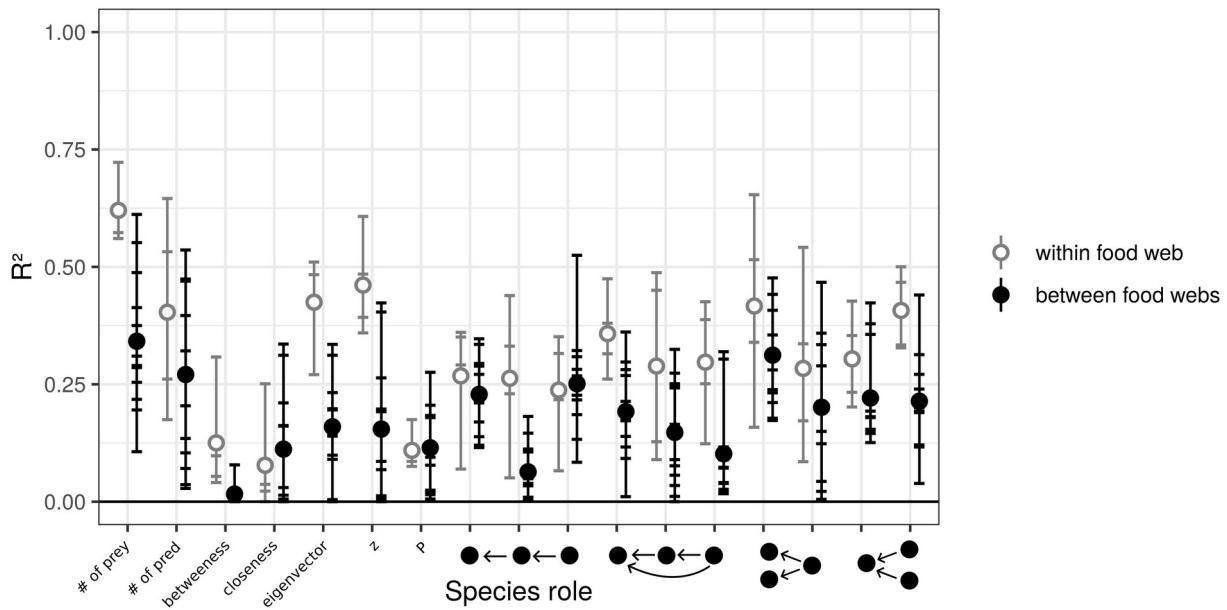
**Figure 5: Trophic interactions model transferability analysis workflow.** Panel (a) shows an example workflow for between-food web predictions. We trained a trophic interaction model using each food web considered in this study (panel a.I). We measured the performance of the four models on a validation subset for within-food web predictions, and the entire food webs for between-food web predictions (panel a.II). Panel (b) maps the spatial domain of each food web. The curves present the pairs of food web model used to make between-food web predictions. Within-food web predictions are not shown in the figure. Bar plots illustrate the proportion of amphibians (green), birds (yellow), mammals (purple), and reptiles (red) in each food web.



**Figure 6: Transferability of predictive models across geographic, environmental and phylogenetic distance.** In each panel, the points present the predictive performance (as area under the receiver operating curve, AUC) of a model food web prediction combination (16 combination; Table 2). For each panel, the thick line is the median effects while the envelope is the 95% credible interval on predictive performance built using the posterior predictive distribution of geographic, environmental and phylogenetic distances. Panel (a) illustrates the total (includes the indirect effects of environmental and phylogenetic distances; turquoise) and partial (controlling for environmental and phylogenetic distances; dark blue) effects of geographic distance. Panel (b) shows the effect of environmental distance after controlling for geographic distance. Panel (c) presents the effect of phylogenetic distance after controlling for geographic distance.

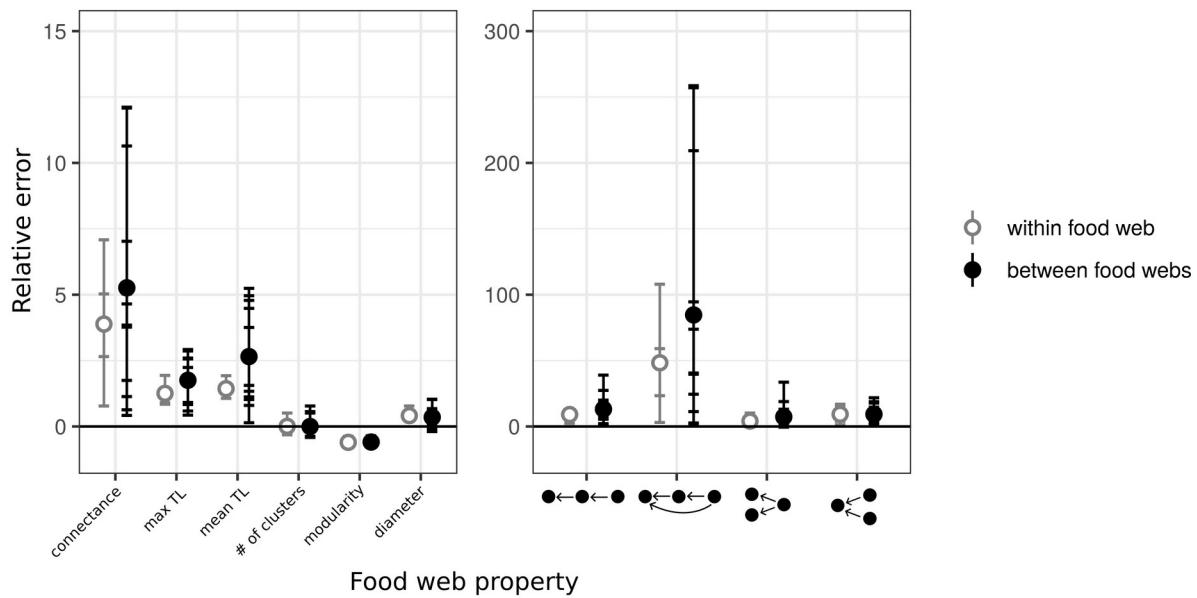


**Figure 7: Predicting species interactions.** Model predictive performance for the species interactions in alternative food webs. Each point shows the model performance in predicting the prey and predators of a single species using area under the receiver operating curve (AUC) as a metric. For each panel, the thick line is the median effect while the envelope is the 95% credible interval on predictive performance built using the posterior predictive distribution of (a) distance to the nearest taxon, (b) mean functional pairwise distance and (c) normalized degree on predictive performance.



**Figure 8: Predicting species functional role.** Performance of models to predict species functional role measured as the proportion of the variance in trophic positions explained by the models ( $R^2$ ). From left to right, the figure shows the number of prey, number of predators, betweenness, closeness, eigenvector centrality, within-module degree (z), participation coefficient (P), times the species is at a specific position in the linear chain (predator – consumer – resource), intraguild predation (omnivore – consumer – resource), direct competition (2 consumers – 1 resource), and apparent competition motifs (1 consumer – 2 resources). The short horizontal lines highlight the  $R^2$  for each role, model, and food web predicted combination. Grey open dots are the mean  $R^2$  for within-food web predictions. Full black dots are the mean  $R^2$  for between-food web predictions.





**Figure 9: Prediction error of global food web properties.** Relative error is the difference between the predicted and the empirical estimates divided by the empirical estimate. From left to right, the figure shows the relative error for connectance, maximum trophic level, mean trophic level, number of clusters, modularity, diameter, number of linear chain (predator – consumer – resource), intraguild predation (omnivore – consumer - resource), direct competition (2 consumers – 1 resource), and apparent competition (1 consumer – 2 resources) motifs. Short horizontal lines are the relative errors for each property, model, and food web predicted combination. Grey open dots are the mean relative errors for within - food web predictions and full black dots are the mean relative errors for between -food web predictions. We divided the figure in two panels because the food web properties in panels (a) and (b) had widely different relative errors.

## Bridging chapter 2 and chapter 3

In chapter 1 and 2, I investigated how functional traits influence the Eltonian niche of terrestrial vertebrates (i.e., relationship to resource and enemies). I showed that these trait-interaction relationships remain relatively constant across taxa and ecosystems given enough environmental and evolutionary similarities. In these two chapters, I assumed that the niche space (environment and interaction spaces) occupied by species do not change in time. However, we know that temporal variations, such as seasonal regimes, in predation, resource availability and climate are ubiquitous in nature. Therefore, species are ‘moving’ within their niche space, shifting resources, enemies, and climatic conditions. To minimize this ‘niche movement’ that requires flexibility, some species migrate, creating a energetic trade-off between within niche movement and movement in geographic space.

In chapter 3, I explore this trade-off in North American birds. I focus on birds, taking advantage of the large amount of information available for this group. I also focus on movement in climatic space (Grinnellian niche) due to the lack of knowledge on seasonal diet for most species. I investigate how traits influence the trade-off between within niche and geographic movements using weekly distribution data, daily climatic variables, and species-level traits. I also ask how these trait-migrating strategy relationships drive seasonal variations in two well-studied biogeographical patterns: the decline in diversity towards the pole (latitudinal diversity gradient) and the increase in body mass with decreasing temperature (Bergmann’s rule).

# **Chapter 3: Consequences of seasonal geographic and niche movements for North American bird biogeography**

**Dominique Caron<sup>1,2</sup>, Maximiliane Jousse<sup>1,2</sup>, Morgan W. Tingley<sup>3</sup>, Dirk Nikolaus Karger<sup>4</sup>,**

**Dominique Gravel<sup>2,5</sup>, Laura J. Pollock<sup>1,2</sup>**

## **Affiliations:**

<sup>1</sup> Department of Biology, McGill University, Montreal, QC, Canada

<sup>2</sup> Quebec Centre for Biodiversity Sciences, Montreal, QC, Canada

<sup>3</sup> Department of Ecology and Evolutionary Biology, University of California, Los Angeles, United States

<sup>4</sup> Swiss Federal Research Institute WSL, Birmensdorf, Switzerland

<sup>5</sup> Département de biologie, Université de Sherbrooke, Sherbrooke, QC, Canada

The following is in preparation for submission to *Global Ecology and Biogeography*

## **Abstract**

Seasonality plays an important, yet often overlooked role in explaining and predicting ecological processes and distributions. Birds respond to seasonal variation through a well-known energetic trade-off of enduring winter or escaping it via seasonal migration. Less is known about how traits explain this trade-off and the degree to which these strategies impact classic ecogeographical relationships. We address these questions by asking: (1) which traits (body mass, hand-wing index, and diet) are most closely associated with the strategy of enduring winter ('niche-switching'; movement within climatic space) versus escaping it ('niche-tracking'; movement in geographic space) and (2) how species with different strategies influence the latitudinal diversity gradient and the assemblage-level temperature-size rule. We quantify the degree of 'niche-tracking' versus 'niche-switching' for 546 North American bird species using weekly distribution estimates and daily climate data at a fine spatial scale (2.8 x 2.8 km grid) and model the degree of 'niche-tracking' as a function of species traits. We find a non-linear relationship between species movement in geographic space and in climatic space, with species migrating longer distances minimizing movement within their niche until a certain distance (4 5000 km). Very long distance migrants seem to track other factors (e.g., resources) as they tend to track less climatic conditions. We also found that species traits are strongly tied to movements both in geographic and climatic spaces. Birds only follow the classic richness gradient during winter, but not in summer driven by the seasonal movement of 'niche-trackers'. Finally, we find contrasting effects of season on the body mass-temperature relationship among 'niche-trackers' and 'niche-switchers', explaining some of the discrepancies in the literature. In the current context of changing climate and shifting seasonal dynamics, predictions of biodiversity responses without considering seasonality would be, at best, part of the story and, at worst, misleading.

## Introduction

Every year, almost 20% of the world's bird species migrate, sometimes over thousands of kilometres, in response to seasonal variations in the environment (Kirby et al., 2008). This extraordinary adaptation to seasonality presumably evolved to maintain fidelity to breeding grounds, while escaping harsh conditions (Winger et al., 2019). Migration has the advantage over other adaptations to seasonal environments (e.g., hibernation, freeze tolerance) in that migrating species can maintain similar thermal conditions throughout the year, but migration also brings movement costs (Gómez et al., 2016; Nakazawa et al., 2004). Indeed, Somveille et al. (2015, 2019) demonstrated that breeding and wintering destinations of migratory birds are optimized to avoid unfavorable climatic conditions, competition, and minimize geographic distance travelled. Therefore, evidence suggests there is a fundamental trade-off between the costs of movement within geographic space and within climatic niche space (Gómez et al., 2016; Nakazawa et al., 2004), but it is poorly understood how the differences between species change the balance of this trade-off.

The degree to which species minimize movement within their climatic niche space (hereinafter referred to as 'niche-tracking' *sensu* Nakazawa et al., 2004), as oppose to sustaining large variation in climatic conditions (hereinafter referred to as 'niche-switching' *sensu* Nakazawa et al., 2004), ultimately involves geographic, morphological, life history, predatory, and feeding habit considerations (J. Cohen & Jetz, 2022; McKinnon et al., 2010; Zurell et al., 2018). For example, Zurell et al. (2018) studied the overlap in breeding and wintering climatic conditions experienced by 717 Holarctic species and found that region, range position, and range size were important determinants of tracking. They also found that larger species tended to experience more similar breeding and wintering climatic conditions, contrary to what we might expect and to the recent findings of J. Cohen & Jetz (2022). Larger species

typically need more energy to migrate (Watanabe, 2016) and can tolerate colder temperatures (Blackburn et al., 1999). Therefore, it should be relatively more beneficial for bigger species to sustain variation in climatic conditions, resulting in an overall broader niche, than to closely track their niche optimum throughout the year. In contrast, species with larger hand-wing-index, a measure of flight efficiency (Sheard et al., 2020), should favour ‘niche-tracking’ since efficient fliers can migrate longer distance at lower energetic cost. ‘Niche-switching’ should also be associated with species with flexible diets. For example, many birds, like the Northern cardinal (*Cardinalis cardinalis*), seasonally switch their diet, foraging in the canopy to eat insects and berries during summer months and on the ground for grains the rest of the year (Gill, 2007). While these examples show how certain traits explain why some birds favour movement within climatic space over movement in geographic space, we are still lacking evidence for other traits and how they combine to explain the trade-off.

If traits can explain the ‘niche-tracking/switching’ trade-off, this would lead to a more mechanistic understanding of the large seasonal shifts in diversity increasingly highlighted in recent literature. For example, Ng et al. (2022) recently found seasonal shifts in biomass, abundance, and richness of nocturnally migrating landbirds in the United States. They found that birds are more abundant and more diverse in the Southeast and the Pacific coast during winter, whereas the Northeastern and Midwestern United States are home to more individuals and species during summer. Similarly, La Sorte et al. (2022) found that, during migration, bird richness was highest on islands within the northern mid-latitudes, whereas, during winter, bird richness was highest on islands within tropical latitudes. The trait composition of bird assemblages also vary across seasons. For example, Jarzyna & Stagge (2023) recently found that in the Northeastern United States, functional richness is highest in winter despite a loss in species richness. Together these results suggest a strong functional basis for seasonal diversity patterns, but the extent to which different traits drive major diversity gradients is less understood.

One such gradient of longstanding interest is the relationship between body mass and latitude. Originally proposed by Bergmann (1848), the rule is traditionally defined as ‘races of warm blooded vertebrates from cooler climates tend to be larger than races of the same species from warmer climates’ (Mayr, 1956). It has now been extended to the relationship between size and latitude, for ectotherms, and across closely related species (as was originally proposed by Bergmann; Blackburn et al., 1999). Indeed, this relationship has strong support in many taxa, including birds (Blackburn & Gaston, 1996; He et al., 2023; Meiri & Dayan, 2003; Youngflesh et al., 2022). However, this prediction is potentially diluted by migration, because the selective pressure for increased body size with latitude should be weaker for birds that escape the harsh conditions of winter (‘niche-trackers’). There are mixed results for the few studies that have tested this expectation. In support, Ramirez et al. (2008) found a stronger body size gradient across latitudes during winter for New World birds. Studies have also found stronger support for Bergmann’s rule for sedentary birds than for migratory birds within (Meiri & Dayan, 2003) and across species (Mainwaring & Street, 2021), but another study found the opposite trend (He et al., 2023). Bergman’s rule is further complicated by recent findings that smaller birds tend to migrate further and track climate more than larger birds (J. Cohen & Jetz, 2022; Soriano-Redondo et al., 2020). Consequently, it is still unclear how the relationship between body mass and migratory behaviour impacts Bergmann’s rule across seasons.

To better understand how seasonality influences Bergman’s rule and the latitudinal biodiversity gradient more generally, we need to first link ‘niche-tracking’ and ‘switching’ to traits and then evaluate biodiversity patterns in this context. So far, the study of bird ‘niche-tracking’ and biogeography has remained largely disconnected, likely, in part, due to the necessity for large spatial scale and high temporal resolution distribution data for entire bird communities and climate variables. Recent data

products make this challenge more possible with weekly bird distributions, daily climate data, and species-specific traits. Now, we can better explore how the relationship between species migration and the degree of ‘niche-tracking’ shapes the distribution of bird diversity and traits across seasons at macroecological scales.

Our objectives are to determine (1) how body mass, hand-wing index, and trophic guild influence the trade-off between movement within climate niche space and movement in geographic space, and (2) the consequences of those relationships on two ecogeographical rules: the latitudinal diversity gradient and Bergmann’s rule. To do this, we combine weekly estimates of species abundances and occurrences for most of the bird diversity in the United States and Canada, along with daily climatic variables, and species-level traits and address the following hypotheses. First, we expect a trade-off between ‘niche-tracking’ and ‘switching’, and for that trade-off to be explained by traits. We expect to see this trade-off with a negative relationship between the distance travelled by species in geographic space and in climatic space (H1). For traits (H2), we expect larger species to favor movement in climate space and species with a larger hand-wing index (more flight efficiency) to favor movement in geographic space. We also expect guilds feeding on seasonal resources (e.g., invertivore, frugivore, nectarivores) to minimize movement in climatic space compared to guilds feeding on less seasonal resources (e.g., omnivores and granivores). Second, we expect ‘niche-trackers’ to dilute or negate classic biodiversity gradients. Specifically, we expect (H3) that the latitudinal richness gradient is lessened in the summer, when ‘niche-trackers’ migrate northward, and (H4) an increase in average body mass toward the poles (Bergmann’s rule) only for ‘niche-switchers’ and not ‘niche-trackers’ who escape the body mass constraints imposed by harsh winters.

## **Material and methods**

### *Bird distributions and traits*

We extracted weekly estimates of bird occurrence and relative abundance from the 2022 version of eBird Status and Trends (Fink et al., 2020, 2023) for bird species in the Avibase (Lepage et al., 2014) checklists for Canada and United States. eBird Status and Trends relates eBird observations between 2008 and 2022 to elevation, topography, and land cover in a hierarchical machine learning model that takes into account observation effort and temporal scales to model species distribution for each week of the year. We extracted the data using the R package ebirdst (Strimas-Mackey et al., 2022) at the highest resolution available ( $2.8 \times 2.8$  km) within continental North and South America. Predictions for all 52 weeks within the study area were available for 546 species out of the 824 birds in Avibase checklists.

We extracted the body mass, hand-wing index, and trophic guild for the 546 species from AVONET (Tobias et al., 2022). AVONET compiles species-level trait means for all the world's bird species. Hand-wing index is a quantitative morphological trait correlated to flight efficiency (Sheard et al., 2020). Trophic guild is a categorical trait describing species major resource types: frugivore (6 species), granivore (40 species), nectarivore (14 species), terrestrial herbivore (11 species), aquatic herbivore (16 species), invertivore (203 species), omnivore (112 species), vertivore (32 species), aquatic predator (110 species), and scavenger (2 species).

### *Climate data and tracking*

To match climate data to bird distributions, we extracted daily minimum near-surface air-temperature and daily precipitation rates between 2008 and 2022 from CHELSA v.2.1 (Karger et al., 2017, 2021) resampled to the 2.8 km grid of eBird Status and Trends within continental North and South America

by using the mean of all 1 km grid cells that overlapped with a 2.8 km grid cell. We use daily minimum near-surface air-temperature and precipitation rates because they have been shown to greatly control species richness and distributions (Qian, 2010) and are usually not highly collinear. We calculated weekly climatological means by averaging daily minimum near-surface air-temperatures and daily precipitation rates for the years 2008 to 2022 for each week of the year, starting January 1<sup>st</sup> and ending December 31<sup>st</sup>.

We calculated the climatic and geographic centroids of each species in each week by estimating the weekly mean latitude, mean longitude, mean daily minimum near-surface air-temperature, and mean daily precipitation rates, weighted by species' abundances within each 2.8 pixel. Before calculating climatic centroids, we centred and scaled weekly mean daily minimum near-surface air-temperature and weekly mean daily precipitation rates to account for different units. Since we are interested in annual movements, we reduced the 'noise' in the geographic and climatic centroids time series caused by week-to-week variations in distributions by taking the moving window average. We used a window size of 5 weeks (e.g., the geographic centroid of a species for week 3 is the average of the centroids of week 1-5), which allowed to smooth the time series without loosing too much of the temporal resolution (Appendix S2). We estimated the distance travelled in climatic space (i.e., degree of 'niche-switching') as the sum of Euclidean distances between the climate centroids of consecutive weeks. Similarly, we estimated distance travel in geographic space as the sum of distances between the geographic centroids occupied by species during consecutive weeks.

#### *Climate condition tracking models*

We first evaluated the relationships between distance travelled in geographic space, distance within climatic space and species traits. First, we modelled distance travelled in climatic space (degree of

‘niche-switching’) as a function of geographic distance for all species. We use a piecewise linear model (segmented regression) as we expect different relationships for long-distance migrators. Somveille et al. (2019) showed that species that migrate intermediate distance tend to track better temperature, whereas species travelling longer distances track available resources. Additionally, the model coefficients (e.g., slopes, breakpoint) of the segmented regression allow direct inference of the biology and behavior of species. Second, we modelled distance travelled in climatic space as a function of body mass, hand-wing index, and trophic guild as a single multiple linear regression. We also fit three separate simple regressions to test the total effect of each trait (i.e., without controlling for the effects of the other traits). Third, we modelled geographic distance travelled as a function of body mass, hand-wing index, and trophic guild, again, as a single multiple linear regression, and as three separate simple regressions. For every model, we log-transformed geographic distance, climatic distance, and body mass to achieve linearity. We also centred and scaled body mass and hand-wing index to account for different units. As these traits are phylogenetically not independent, with closer related species usually showing similar traits, we additionally refitted each model after controlling for phylogenetic relatedness (Appendix S3) by adding a varying intercept over species constrained with a covariance matrix built from the avian phylogeny from Jetz et al. (2012).

Model parameters were estimated with Hamiltonian Monte Carlo (Neal, 2011). We ran four chains, each with 1000 warm-up iterations, followed by 1000 iterations for inference. We diagnosed convergence and adequacy with trace rank plots, posterior predictive checks, and we calculated the rank-normalized potential scale reduction factor on split chains for all runs (Vehtari et al., 2021; Appendix S4). We conducted the analyses using Stan (Carpenter et al., 2017) through the package brms in R (Bürkner, 2018).

### *Seasonality of ecogeographical rules*

We estimated the species richness for each pixel and week by summing probabilities of occurrence across all eBird Status and Trends layers. We calculated total bird richness, richness for each trophic guild, richness of ‘niche-trackers’, and richness of ‘niche-switchers’. Although ‘niche-switching’ and ‘tracking’ is a gradient, we treated it here as a binary variable as a convenient simplification to explore the role of species that travel less distance in climatic space (‘niche-trackers’) than average on the seasonal variation of the ecogeographical rules. We defined ‘niche-switchers’ as species that travels more within their climatic niche than the average species. Similarly, we defined ‘niche-trackers’ as species that travels less within their climatic niche than the average species. We compared the diversity distributions across latitude for all 52 weeks, and additionally also only between the summer (from June 18 to September 17) and winter seasons (from December 18 to March 18). We fit a generalized additive model using a random sample of 10 000 pixels (0.5% of all pixels) to reduce spatial autocorrelation effects with species richness as response variable and latitude as predictor using the R package mgcv (Wood, 2011).

We used an assemblage-based approach to explore the seasonal variation in Bergmann’s rule at the community level (Gaston et al., 2008; Olalla-Tárraga et al., 2010). For each pixel and week, we calculated community weighted mean body mass:

$$\log(CWM_{BM_{jk}}) = \frac{\sum_{i=1}^N p_{ijk} \times \log(BM_i)}{\sum_{i=1}^N p_{ijk}},$$

where  $CWM_{BM_{jk}}$  is the community weighted mean body mass for pixel  $j$  on week  $k$ ,  $p_{ijk}$  is the probability of occurrence of species  $i$  in pixel  $j$  on week  $k$ , and  $BM_i$  is the body mass of species  $i$ . We used probability of occurrences and not relative abundances because eBird Status and Trends’ relative abundance estimates are comparable for a given species but not across species. Similar to richness, we

calculated the average body mass across all species, for each trophic guild, and for ‘niche-trackers’ and ‘niche-switchers’. We fit a linear model for summer and winter with log-transformed body mass as response and the latitude as predictor with varying slopes for trophic guilds and for strategy (‘niche-trackers’ vs. ‘niche-switchers’). We compared parameter estimates (slopes) associated to evaluate whether adherence to Bergmann’s rule across species differed between seasons, guilds, and strategies. Linear models were fit using a random sample of 10 000 pixels (0.5% of all pixels) and were estimated with Hamiltonian Monte Carlo (Neal, 2011) through Stan and the R package brms.

## Results

### *Climate condition tracking*

Long-distance migrants travelled less distance within their climatic niche more than residents and short distance migrants (Figure 1). Climatic distance travelled by a bird species decreased with distance travelled in geographic space (posterior mean slope = -0.09; 95% CrI = [-0.12, -0.05]) until a breakpoint at a geographic distance of 4457 km [95% CrI = [3448 km, 5726 km], after which climatic distance increased with increasing geographic distance (posterior mean slope = 0.26; 95% CrI = [0.19, 0.34]). Distance travelled in geographic space explained 15.1% ( $R^2$ ; 95% CrI = [10.3%, 20.3%]) of the variation in climatic distance travelled by species.

Body mass, hand-wing index, and trophic guild explained 38.8% ( $R^2$ ; 95% CrI = [33.7%, 43.8%]) of the variation in migration distance travelled across species (Table 1a). In all guilds, birds with large hand-wing indices migrated further than birds with small hand-wing index. Across guilds, larger species travelled longer distances. However, after controlling for hand-wing index, smaller species tended to travel longer distances compared to bigger birds within most guilds. Aquatic predators tended

to travel more than most other guilds, whereas frugivores tended to travel less distances. These guild effects were generally weaker after controlling for morphological differences.

Body mass, hand-wing index, and trophic guild of species explained 16.1% ( $R^2$ ; 95% CrI = [10.8%, 21.5%]) of the variation in climatic distance travelled across species (Table 1b). Across guilds, larger birds and birds with smaller hand-wing index tracked more climatic conditions across seasons. The effect of body mass remained relatively stable after controlling for differences in hand-wing index and guilds, whereas the effect of hand-wing index was stronger after controlling for other traits within many guilds. We found no significant difference among guilds in how much they track climate before or after controlling for morphological differences.

Overall, contrary to our expectations, these results suggest that species with larger body mass tend to have a larger hand-wing index and diet that allow them to migrate further to minimize movement in climatic space. Interestingly, species with larger hand-wing index travel longer distances both in geographic and climatic spaces. Finally, we found little evidence that guilds feeding on more seasonal resources (invertivores, aquatic predators, nectarivores) migrate longer distances to minimize movement within their climatic niche.

#### *Seasonality of ecogeographical rules*

During winter weeks, bird richness followed the expected latitudinal diversity gradient with a decrease of diversity toward northern latitudes (Figure 3). However, the relationship changed during summer weeks, with an increase in richness from south to mid-latitude (~40°), followed by a decrease towards higher latitudes. This observation is explained by the movement of ‘niche-trackers’ from south to north during spring migration. ‘Niche switchers’ did not follow a decrease in richness in both summer and

winter. In northern latitudes, the richness of ‘niche-trackers’ became greater than ‘niche-switchers’ during winter weeks, whereas in summer and in southern latitudes, ‘niche-trackers’ represented the majority of bird richness. Based on Birdlife range maps of all species (Birdlife International, 2022), it does not seem like incomplete sampling of the north American bird assemblage (eBird status and trends modelled 546 species out of the 824 birds in Avibase checklists) would impact our findings (Appendix S1).

Across all bird species, we did not find evidence for Bergmann’s rule: increased average body mass across species at lower temperature and higher latitudes (Figure 4). However, within each niche strategy (‘niche-tracking’ and ‘niche-switchers’), we found evidence for Bergmann’s rule in both seasons. Interestingly, we found a stronger relationship for ‘niche-trackers’ in winter, but the opposite for ‘niche-switcher’, with average body mass more strongly correlated with latitude and temperature in summer. Overall, we found little evidence for Bergmann’s rule within most trophic guilds neither in summer, nor in winter. The only guilds that adhered to the rule were omnivores and aquatic herbivores both during summer and winter. We found the opposite relationship in granivores and invertivores in both seasons. For terrestrial herbivores and aquatic predators, we found evidence for Bergmann’s rule in winter, but not in summer. Finally, vertivores tended to be bigger at lower latitudes and higher temperature in summer, but this relationship faded in winter.

## Discussion

We asked whether the trade-off in distance travelled in climate versus geographic space can be explained by traits, and whether ecogeographical patterns are strongly influenced. We found evidence to support both. First, we observed evidence of a trade-off, as movement within climate niche

decreased with distance travelled in geographic space, up to approximately 4500 km, beyond which climate niche movement increases with geographic distance. This trade-off was attributed to larger birds with better migration capacities migrating further and better tracking climate conditions across seasons. Second, we found that seasonal migration strongly influenced both the latitudinal richness gradient and Bergmann's rule. The latitudinal richness gradient was clearly visible in winter with a decline towards the pole, but unimodal in summer when 'niche-trackers' moved north. Similarly, while Bergmann's rule was not observed year-round across all birds or within most trophic guilds, 'niche-switchers' and 'niche-trackers' adhered to Bergmann's rule, but with different seasonal influences.

As expected, species that travel longer distances in geographic space tend to travel less within their climatic niche supporting the hypothesis that migrating birds seek areas with similar climate throughout the year (Somveille et al., 2019; Winger et al., 2019). Although the trade-off between energy spent for movement in geographic space and in climatic space stands on solid theoretical grounds, the empirical evidence in the literature supporting it has received mixed support. Many previous studies found some evidence of 'niche-tracking' for migratory species (but see Dufour et al., 2020 and Ponti et al., 2020) when compared to null expectations (Eyres et al., 2020; Zurell et al., 2018) or resident species (Gómez et al., 2016), but many found no (Gómez et al., 2016; Laube et al., 2015) or a negative (Ponti et al., 2020; Zurell et al., 2018) relationship between migration distance and 'niche-tracking'. In contrast, similar to our results, Cohen & Jetz (2022) found a positive relationship between migration distance and 'niche-tracking'. Differences in methodology, data, and study systems could explain the mix of results across studies. For example, many studies used coarse breeding and wintering range maps (Ponti et al., 2020; Eyres et al., 2020; Zurrell et al., 2018; Dufour et al., 2020; Laube et al., 2015) and uses yearly or monthly climatic averages (e.g., Laube et al., 2015; Gómez et al., 2016; Dufour et al., 2020; Eyres et al., 2020; Ponti et al., 2020; Zurrell et al., 2018), missing potentially important

variations in the climate-occurrence relationships at both finer spatial and temporal resolution. Here, we used fine-grained distributions for each week together with daily climatic data. Finally, we found that the relationship between migration distances and climatic distances is segmented, with a positive relationship for very long-distance migrants. This finding parallels the results by Somveille et al (2019) who suggested that long-distance migrants prioritize areas with enhanced resource accessibility rather than closely tracking thermal conditions. Here, we only considered movement within the Grinnellian (environment) dimensions of the niche. Many species switch or track their resources (Eltonian niche), which could represent another type of energetic trade-off. It would be interesting to explore and compare how species shift different dimensions of their niche across seasons.

We also found support the hypothesis that traits influence the trade-off between the cost of movement and the cost of adapting to broader climatic conditions (Somveille et al., 2015, 2019), but that these trait-movements relationships are more complex than expected. Birds with larger hand-wing index, indicative of flight efficiency, tended to travel longer distances in both geographic and within climate niche spaces. This result partly diverge from our initial expectation that more efficient migrators would travel longer distances to minimize movement within their niche. This discrepancy might be attributed to the observation that very long-distance migrants, likely possessing a large hand-wing index, track resource availability rather than closely tracking climate conditions (Somveille et al., 2019). Additionally, species with greater migration capabilities may inhabit more seasonal areas, experiencing substantial climate variations within seasons despite evading winters. This hypothesis finds support in our discovery of a negative relationship between hand-wing index and climatic distance when exclusively considering inter-seasonal movements (distance between breeding and non-breeding centroids; Appendix S2). Surprisingly, but in line with the results of Zurell et al. (2018), we also found that, across all birds, bigger species tend to travel longer distances and minimized movements within

their climatic niches. These body mass effects however mainly faded when accounting for variations in movements and hand-wing index across guilds. For instance, aquatic predators, typically characterized by larger body size and hand-wing index, exhibited longer travel distances in geographic space while also minimizing movement within their climatic niche. Other traits and considerations likely influence migration behaviour. For example, Soriano-Redondo et al. (2020) found that life history traits influence migration distance, with migrant birds living faster paces of life (larger clutch size, earlier maturity, and shorter lifespan) than resident species. Also, we assumed that migration energetic cost is linear with distance, but migration routes are much more complex and depends on barriers (Cano et al., 2020; Lok et al., 2015), winds (Kranstauber et al., 2015; Norevik et al., 2020), stop-overs (E. B. Cohen et al., 2021), among other factors. Thus, geographic factors likely influence how species balance the migration-'niche-switching' trade-off (Zurell et al., 2018).

As was showed in previous research (Jarzyna & Stagge, 2023; La Sorte et al., 2022; Ng et al., 2022), we found a profound seasonal shift in the latitudinal diversity gradient of birds across the United States and Canada. In fact, we found the expected latitudinal diversity gradient, species richness peaking towards the equator and decreasing towards the pole during winter. In contrast, we found a unimodal distribution of richness across latitude for the summer, with diversity peaking at mid-latitude, around 40°. This shift in diversity was driven by movement of species that minimize movement within their climatic niche more than the average ('niche-trackers') across seasons. In fact, species that move more within their niche than the average ('niche-switchers') still adhere to the expected latitudinal diversity gradient. Dufour et al. (2024) recently suggested that the inverse latitudinal diversity gradient observed in some bird clades is the result of migratory behaviour allowing the colonization of highly seasonal areas where diversification was increased compared to the ancestral tropical biomes. Our results support this hypothesis, with the classic decrease in bird richness at higher latitudes observed in winter,

when birds occupy their ancestral range, but the inverse pattern in summer, when ‘niche-trackers’ occupy ranges that migration allowed them to colonize. Future work could seek to link the evolution of ‘niche-tracking’ behaviour with diversity gradients.

We also found mixed evidence for interspecific Bergmann’s rule. While we did not observe the classic negative relationship between average body size and latitude across all birds or within most guilds, we did find evidence for Bergmann’s rule within ‘niche-switchers’ and ‘niche-trackers’ throughout the year. Interestingly, the relationship between body size and temperature was stronger in summer for ‘niche-switchers’, but weaker for ‘niche-trackers’. This may indicate that different mechanisms shape Bergmann’s rule within these two groups (Blackburn et al., 1999; Henry et al., 2023). For example, if heat dissipation (larger surface-to-volume ratio helps smaller individuals stay cool in warm and dry environments) is the primary mechanism, we would expect the relationship to be stronger in summer, as observed among ‘niche-switchers’. In contrast, if heat conservation (smaller surface-to-volume ratio minimizes heat loss in cold temperatures) mainly shapes body size-temperature relationships, we would expect a stronger relationship in winter, as observed for ‘niche-trackers’. Further research is needed to distinguish between the role of these and other mechanisms among different migratory strategies. Also, we used an assemblage-based approach to test the temperature-size relationship, but Bergmann’s rule is traditionally a species-level relationship. It is possible that processes acting at the community level predict relationships between body size and temperature opposite to Bergmann’s rule (Classen et al., 2017). For example, in energy-restricted habitats, increased extinction risk of larger animals could result in an overall decline in the mean body sizes within species communities (Aava, 2001; Brown & Maurer, 1989). Therefore, it is possible that, within some guilds or during certain seasons, community-level processes outweigh species-level processes to explain the community mean body size-temperature relationship, but this would need to be tested.

Describing how and why trait distributions change has important implications on the stability and functioning of ecosystems. For example, species body mass and dietary guilds explain many ecological processes such as predation (Brose, 2010; Jacob et al., 2011) and competition (Leyequién et al., 2007; Monterroso et al., 2020). Given the shifts in diversity, body mass, and guild distribution across space and seasons, we predict that species interactions and the networks they encompass will sustain substantial rewiring across seasons. Seasonal forcing has the potential to stabilize or destabilize ecological networks (Donohue & Piiroinen, 2016; Duchenne et al., 2022; Saavedra et al., 2016). Of course, here we only indirectly investigate variation in food webs using two trophic traits, guild and body mass, but directly using interaction data would enable us to tackle the question of seasonal variation of food webs at continental scales. Our results add to the growing evidence that it is imperative to integrate seasonality in ecological models to describe, predict, and understand populations, communities, and ecosystems (White & Hastings, 2020). This is especially critical given that we know migratory species are declining (Bairlein, 2016), and that the length, intensity, and timing of seasons are changing globally (Stevenson et al., 2015).

## References

- Aava, B. (2001). Primary productivity can affect mammalian body size frequency distributions. *Oikos*, 93(2), 205–212. <https://doi.org/10.1034/j.1600-0706.2001.930204.x>
- Bairlein, F. (2016). Migratory birds under threat. *Science*, 354(6312), 547–548. <https://doi.org/10.1126/science.aah6647>
- Bergmann, C. (1848). *Über die Verhältnisse der Wärmeökonomie der Thiere zu ihrer Grösse*. Vandenhoeck und Ruprecht.
- Blackburn, T. M., & Gaston, K. J. (1996). Spatial Patterns in the Body Sizes of Bird Species in the New World. *Oikos*, 77(3), 436–446. <https://doi.org/10.2307/3545933>
- Blackburn, T. M., Gaston, K. J., & Loder, N. (1999). Geographic gradients in body size: A clarification of Bergmann's rule. *Diversity and Distributions*, 5(4), 165–174. <https://doi.org/10.1046/j.1472-4642.1999.00046.x>
- Brose, U. (2010). Body-mass constraints on foraging behaviour determine population and food-web dynamics. *Functional Ecology*, 24(1), 28–34. <https://doi.org/10.1111/j.1365-2435.2009.01618.x>
- Brown, J. H., & Maurer, B. A. (1989). Macroecology: The Division of Food and Space Among Species on Continents. *Science*, 243(4895), 1145–1150. <https://doi.org/10.1126/science.243.4895.1145>
- Bürkner, P.-C. (2018). Advanced Bayesian Multilevel Modeling with the R Package brms. *The R Journal*, 10(1), 395. <https://doi.org/10.32614/RJ-2018-017>
- Cano, N., Bayly, N. J., & Wilson, S. (2020). Is there more than one way to cross the Caribbean Sea? Migratory strategies of Nearctic-Neotropical landbirds departing from northern Colombia. *Journal of Avian Biology*, 51(7). <https://doi.org/10.1111/jav.02394>

- Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., & Riddell, A. (2017). *Stan: A Probabilistic Programming Language*. *Journal of Statistical Software*, 76(1). <https://doi.org/10.18637/jss.v076.i01>
- Classen, A., Steffan-Dewenter, I., Kindeketa, W. J., & Peters, M. K. (2017). Integrating intraspecific variation in community ecology unifies theories on body size shifts along climatic gradients. *Functional Ecology*, 31(3), 768–777. <https://doi.org/10.1111/1365-2435.12786>
- Cohen, E. B., Horton, K. G., Marra, P. P., Clipp, H. L., Farnsworth, A., Smolinsky, J. A., Sheldon, D., & Buler, J. J. (2021). A place to land: Spatiotemporal drivers of stopover habitat use by migrating birds. *Ecology Letters*, 24(1), 38–49. <https://doi.org/10.1111/ele.13618>
- Cohen, J., & Jetz, W. (2022). *Diverse strategies for tracking seasonal environmental niches at hemispheric scale* [Preprint]. *Ecology*. <https://doi.org/10.1101/2022.07.15.500241>
- Donohue, J. G., & Piironen, P. T. (2016). The effects of predation on seasonally migrating populations. *Theoretical Ecology*, 9(4), 487–499. <https://doi.org/10.1007/s12080-016-0304-1>
- Duchenne, F., Wüest, R. O., & Graham, C. H. (2022). Seasonal structure of interactions enhances multidimensional stability of mutualistic networks. *Proceedings of the Royal Society B: Biological Sciences*, 289(1982), 20220064. <https://doi.org/10.1098/rspb.2022.0064>
- Dufour, P., Crochet, P.-A., Condamine, F. L., & Lavergne, S. (2024). Seasonal migration and the evolution of an inverse latitudinal diversity gradient in shorebirds. *Global Ecology and Biogeography*, 33(5), e13817. <https://doi.org/10.1111/geb.13817>
- Dufour, P., Descamps, S., Chantepie, S., Renaud, J., Guéguen, M., Schiffers, K., Thuiller, W., & Lavergne, S. (2020). Reconstructing the geographic and climatic origins of long-distance bird migrations. *Journal of Biogeography*, 47(1), 155–166. <https://doi.org/10.1111/jbi.13700>

Eyres, A., Böhning-Gaese, K., Orme, C. D. L., Rahbek, C., & Fritz, S. A. (2020). A tale of two seasons:

The link between seasonal migration and climatic niches in passerine birds. *Ecology and Evolution*, 10(21), 11983–11997. <https://doi.org/10.1002/ece3.6729>

Fink, D., Auer, T., Johnston, A., Ruiz-Gutierrez, V., Hochachka, W. M., & Kelling, S. (2020). Modeling avian full annual cycle distribution and population trends with citizen science data. *Ecological Applications*, 30(3). <https://doi.org/10.1002/eap.2056>

Fink, D., Auer, T., Johnston, A., Strimas-Mackey, M., Ligocki, S., Robinson, O., Hochachka, W., Jaromczyk, L., Crowley, C., Dunham, K., Stillman, A., Davies, I., Rodewald, A., Ruiz-Gutierrez, V., & Wood, C. (2023). *eBird Status and Trends* [dataset]. <https://doi.org/10.2173/ebirdst.2022>

Gaston, K. J., Chown, S. L., & Evans, K. L. (2008). Ecogeographical rules: Elements of a synthesis. *Journal of Biogeography*, 35(3), 483–500. <https://doi.org/10.1111/j.1365-2699.2007.01772.x>

Gill, F. B. (2007). *Ornithology* (3rd ed). W.H. Freeman.

Gómez, C., Tenorio, E. A., Montoya, P., & Cadena, C. D. (2016). Niche-tracking migrants and niche-switching residents: Evolution of climatic niches in New World warblers (Parulidae). *Proceedings of the Royal Society B: Biological Sciences*, 283(1824), 20152458. <https://doi.org/10.1098/rspb.2015.2458>

He, J., Tu, J., Yu, J., & Jiang, H. (2023). A global assessment of Bergmann's rule in mammals and birds. *Global Change Biology*, 29(18), 5199–5210. <https://doi.org/10.1111/gcb.16860>

Henry, E., Santini, L., Huijbregts, M. A. J., & Benítez-López, A. (2023). Unveiling the environmental drivers of intraspecific body size variation in terrestrial vertebrates. *Global Ecology and Biogeography*, 32(2), 267–280. <https://doi.org/10.1111/geb.13621>

Jacob, U., Thierry, A., Brose, U., Arntz, W. E., Berg, S., Brey, T., Fetzer, I., Jonsson, T., Mintenbeck, K., Möllmann, C., Petchey, O. L., Riede, J. O., & Dunne, J. A. (2011). The Role of Body Size in

- Complex Food Webs: A Cold Case. In A. Belgrano (Ed.), *Advances in Ecological Research* (Vol. 45, pp. 181–223). Academic Press. <https://doi.org/10.1016/B978-0-12-386475-8.00005-8>
- Jarzyna, M. A., & Stagge, J. H. (2023). Decoupled spatiotemporal patterns of avian taxonomic and functional diversity. *Current Biology*, 33(6), 1153-1161.e4.  
<https://doi.org/10.1016/j.cub.2023.01.066>
- Jetz, W., Thomas, G. H., Joy, J. B., Hartmann, K., & Mooers, A. O. (2012). The global diversity of birds in space and time. *Nature*, 491(7424), Article 7424. <https://doi.org/10.1038/nature11631>
- Karger, D. N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R. W., Zimmermann, N. E., Linder, H. P., & Kessler, M. (2017). Climatologies at high resolution for the earth's land surface areas. *Scientific Data*, 4(1), Article 1. <https://doi.org/10.1038/sdata.2017.122>
- Karger, D. N., Wilson, A. M., Mahony, C., Zimmermann, N. E., & Jetz, W. (2021). Global daily 1 km land surface precipitation based on cloud cover-informed downscaling. *Scientific Data*, 8(1), Article 1. <https://doi.org/10.1038/s41597-021-01084-6>
- Kirby, J. S., Stattersfield, A. J., Butchart, S. H. M., Evans, M. I., Grimmett, R. F. A., Jones, V. R., O'Sullivan, J., Tucker, G. M., & Newton, I. (2008). Key conservation issues for migratory land- and waterbird species on the world's major flyways. *Bird Conservation International*, 18(S1), S49–S73. <https://doi.org/10.1017/S0959270908000439>
- Kranstauber, B., Weinzierl, R., Wikelski, M., & Safi, K. (2015). Global aerial flyways allow efficient travelling. *Ecology Letters*, 18(12), 1338–1345. <https://doi.org/10.1111/ele.12528>
- La Sorte, F. A., Somveille, M., Dokter, A. M., & Miller, E. T. (2022). Seasonal species richness of birds on the world's islands and its geographical correlates. *Proceedings of the Royal Society B: Biological Sciences*, 289(1980), 20221105. <https://doi.org/10.1098/rspb.2022.1105>

Laube, I., Graham, C. H., & Böhning-Gaese, K. (2015). Niche availability in space and time: Migration

in Sylvia warblers. *Journal of Biogeography*, 42(10), 1896–1906.

<https://doi.org/10.1111/jbi.12565>

Lepage, D., Vaidya, G., & Guralnick, R. (2014). Avibase – a database system for managing and

organizing taxonomic concepts. *ZooKeys*, 420, 117–135.

<https://doi.org/10.3897/zookeys.420.7089>

Leyequién, E., de Boer, W. F., & Cleef, A. (2007). Influence of body size on coexistence of bird

species. *Ecological Research*, 22(5), 735–741. <https://doi.org/10.1007/s11284-006-0311-6>

Lok, T., Overdijk, O., & Piersma, T. (2015). The cost of migration: Spoonbills suffer higher mortality

during trans-Saharan spring migrations only. *Biology Letters*, 11(1), 20140944.

<https://doi.org/10.1098/rsbl.2014.0944>

Mainwaring, M. C., & Street, S. E. (2021). Conformity to Bergmann's rule in birds depends on nest

design and migration. *Ecology and Evolution*, 11(19), 13118–13127.

<https://doi.org/10.1002/ece3.8034>

Mayr, E. (1956). Geographical character gradients and climatic adaptation. *Evolution*, 10(1), 105–108.

McKinnon, L., Smith, P. A., Nol, E., Martin, J. L., Doyle, F. I., Abraham, K. F., Gilchrist, H. G.,

Morrison, R. I. G., & Béty, J. (2010). Lower Predation Risk for Migratory Birds at High

Latitudes. *Science*, 327(5963), 326–327. <https://doi.org/10.1126/science.1183010>

Meiri, S., & Dayan, T. (2003). On the validity of Bergmann's rule. *Journal of Biogeography*, 30(3),

331–351. <https://doi.org/10.1046/j.1365-2699.2003.00837.x>

Monterroso, P., Díaz-Ruiz, F., Lukacs, P. M., Alves, P. C., & Ferreras, P. (2020). Ecological traits and

the spatial structure of competitive coexistence among carnivores. *Ecology*, 101(8), e03059.

<https://doi.org/10.1002/ecy.3059>

- Nakazawa, Y., Peterson, A. T., Martínez-Meyer, E., & Navarro-Sigüenza, A. G. (2004). SEASONAL NICHES OF NEARCTIC-NEOTROPICAL MIGRATORY BIRDS: IMPLICATIONS FOR THE EVOLUTION OF MIGRATION. *The Auk*, 121(2), 610. [https://doi.org/10.1642/0004-8038\(2004\)121\[0610:SNONMB\]2.0.CO;2](https://doi.org/10.1642/0004-8038(2004)121[0610:SNONMB]2.0.CO;2)
- Neal, R. M. (2011). MCMC Using Hamiltonian Dynamics. In *Handbook of Markov Chain Monte Carlo* (pp. 113–162). CRC Press. <https://doi.org/10.1201/b10905-7>
- Ng, W. H., Fink, D., La Sorte, F. A., Auer, T., Hochachka, W. M., Johnston, A., & Dokter, A. M. (2022). Continental-scale biomass redistribution by migratory birds in response to seasonal variation in productivity. *Global Ecology and Biogeography*, 31(4), 727–739. <https://doi.org/10.1111/geb.13460>
- Norevik, G., Åkesson, S., Artois, T., Beenaerts, N., Conway, G., Cresswell, B., Evens, R., Henderson, I., Jiguet, F., & Hedenstrom, A. (2020). Wind-associated detours promote seasonal migratory connectivity in a flapping flying long-distance avian migrant. *Journal of Animal Ecology*, 89(2), 635–646. <https://doi.org/10.1111/1365-2656.13112>
- Olalla-Tárraga, M. Á., Bini, L. M., Diniz-Filho, J. A. F., & Rodríguez, M. Á. (2010). Cross-species and assemblage-based approaches to Bergmann's rule and the biogeography of body size in Plethodon salamanders of eastern North America. *Ecography*, 33(2), 362–368. <https://doi.org/10.1111/j.1600-0587.2010.06244.x>
- Ponti, R., Arcones, A., Ferrer, X., & Vieites, D. R. (2020). Seasonal climatic niches diverge in migratory birds. *Ibis*, 162(2), 318–330. <https://doi.org/10.1111/ibi.12784>
- Qian, H. (2010). Environment–richness relationships for mammals, birds, reptiles, and amphibians at global and regional scales. *Ecological Research*, 25(3), 629–637. <https://doi.org/10.1007/s11284-010-0695-1>

- Ramirez, L., Diniz-Filho, J. A. F., & Hawkins, B. A. (2008). Partitioning phylogenetic and adaptive components of the geographical body-size pattern of New World birds. *Global Ecology and Biogeography*, 17(1), 100–110. <https://doi.org/10.1111/j.1466-8238.2007.00346.x>
- Saavedra, S., Rohr, R. P., Fortuna, M. A., Selva, N., & Bascompte, J. (2016). Seasonal species interactions minimize the impact of species turnover on the likelihood of community persistence. *Ecology*, 97(4), 865–873. <https://doi.org/10.1890/15-1013.1>
- Sheard, C., Neate-Clegg, M. H. C., Alioravainen, N., Jones, S. E. I., Vincent, C., MacGregor, H. E. A., Bregman, T. P., Claramunt, S., & Tobias, J. A. (2020). Ecological drivers of global gradients in avian dispersal inferred from wing morphology. *Nature Communications*, 11(1), Article 1. <https://doi.org/10.1038/s41467-020-16313-6>
- Somveille, M., Manica, A., & Rodrigues, A. S. L. (2019). Where the wild birds go: Explaining the differences in migratory destinations across terrestrial bird species. *Ecography*, 42(2), 225–236. <https://doi.org/10.1111/ecog.03531>
- Somveille, M., Rodrigues, A. S. L., & Manica, A. (2015). Why do birds migrate? A macroecological perspective. *Global Ecology and Biogeography*, 24(6), 664–674. <https://doi.org/10.1111/geb.12298>
- Soriano-Redondo, A., Gutiérrez, J. S., Hodgson, D., & Bearhop, S. (2020). Migrant birds and mammals live faster than residents. *Nature Communications*, 11(1), Article 1. <https://doi.org/10.1038/s41467-020-19256-0>
- Stevenson, T. J., Visser, M. E., Arnold, W., Barrett, P., Biello, S., Dawson, A., Denlinger, D. L., Dominoni, D., Ebling, F. J., Elton, S., Evans, N., Ferguson, H. M., Foster, R. G., Hau, M., Haydon, D. T., Hazlerigg, D. G., Heideman, P., Hopcraft, J. G. C., Jonsson, N. N., ... Helm, B. (2015). Disrupted seasonal biology impacts health, food security and ecosystems. *Proceedings*

- of the Royal Society B: Biological Sciences*, 282(1817), 20151453.  
<https://doi.org/10.1098/rspb.2015.1453>
- Strimas-Mackey, M., Ligocki, S., Auer, T., & Fink, D. (2022). *ebirdst: Tools for loading, plotting, mapping and analysis of eBird Status and Trends data products*. <https://ebird.github.io/ebirdst/>
- Tobias, J. A., Sheard, C., Pigot, A. L., Devenish, A. J. M., Yang, J., Sayol, F., Neate-Clegg, M. H. C., Alioravainen, N., Weeks, T. L., Barber, R. A., Walkden, P. A., MacGregor, H. E. A., Jones, S. E. I., Vincent, C., Phillips, A. G., Marples, N. M., Montaño-Centellas, F. A., Leandro-Silva, V., Claramunt, S., ... Schleuning, M. (2022). AVONET: Morphological, ecological and geographical data for all birds. *Ecology Letters*, 25(3), 581–597.  
<https://doi.org/10.1111/ele.13898>
- Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., & Bürkner, P.-C. (2021). Rank-Normalization, Folding, and Localization: An Improved  $R^*$  for Assessing Convergence of MCMC (with Discussion). *Bayesian Analysis*, 16(2), 667–718. <https://doi.org/10.1214/20-BA1221>
- Watanabe, Y. Y. (2016). Flight mode affects allometry of migration range in birds. *Ecology Letters*, 19(8), 907–914. <https://doi.org/10.1111/ele.12627>
- White, E. R., & Hastings, A. (2020). Seasonality in ecology: Progress and prospects in theory. *Ecological Complexity*, 44, 100867. <https://doi.org/10.1016/j.ecocom.2020.100867>
- Winger, B. M., Auteri, G. G., Pegan, T. M., & Weeks, B. C. (2019). A long winter for the Red Queen: Rethinking the evolution of seasonal migration. *Biological Reviews*, 94(3), 737–752.  
<https://doi.org/10.1111/brv.12476>
- Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society (B)*, 73(1), 3–36.

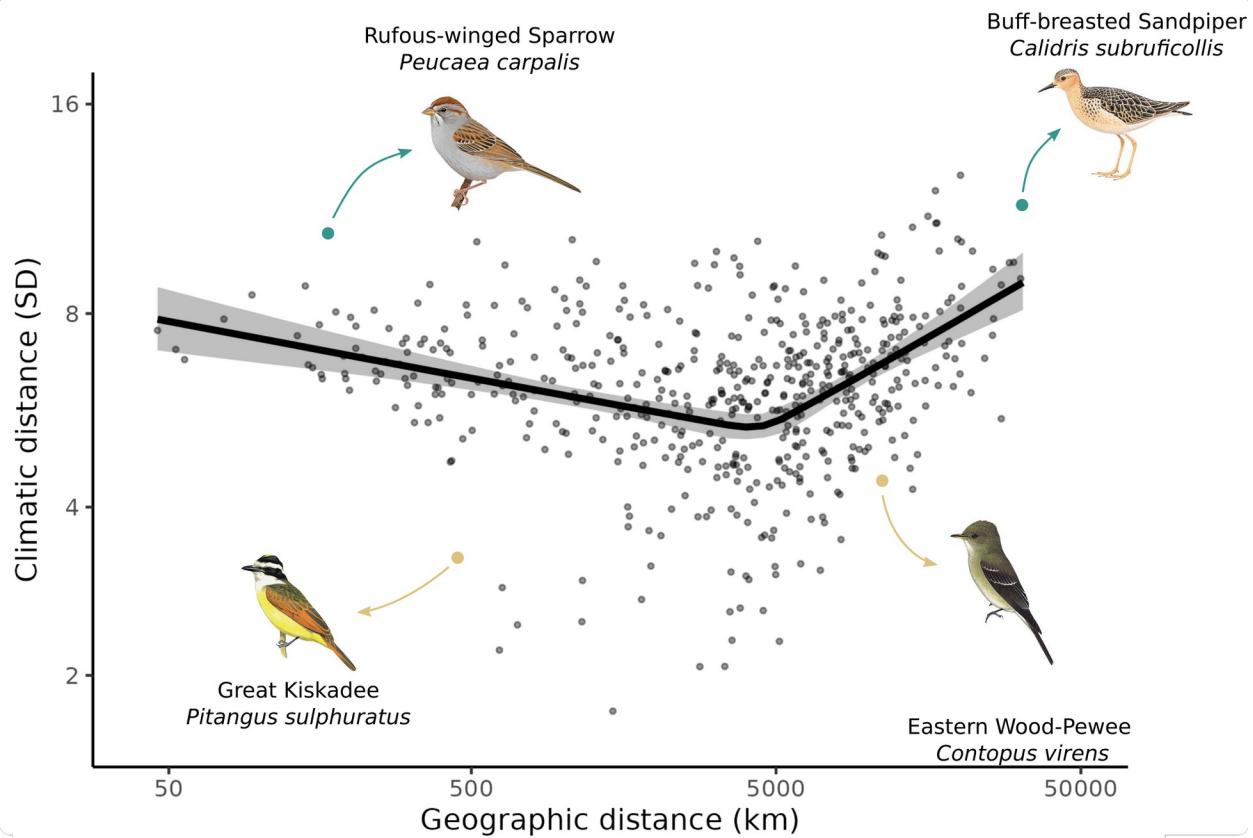
Youngflesh, C., Saracco, J. F., Siegel, R. B., & Tingley, M. W. (2022). Abiotic conditions shape spatial and temporal morphological variation in North American birds. *Nature Ecology & Evolution*, 6(12), Article 12. <https://doi.org/10.1038/s41559-022-01893-x>

Zurell, D., Gallien, L., Graham, C. H., & Zimmermann, N. E. (2018). Do long-distance migratory birds track their niche through seasons? *Journal of Biogeography*, 45(7), 1459–1468.  
<https://doi.org/10.1111/jbi.13351>

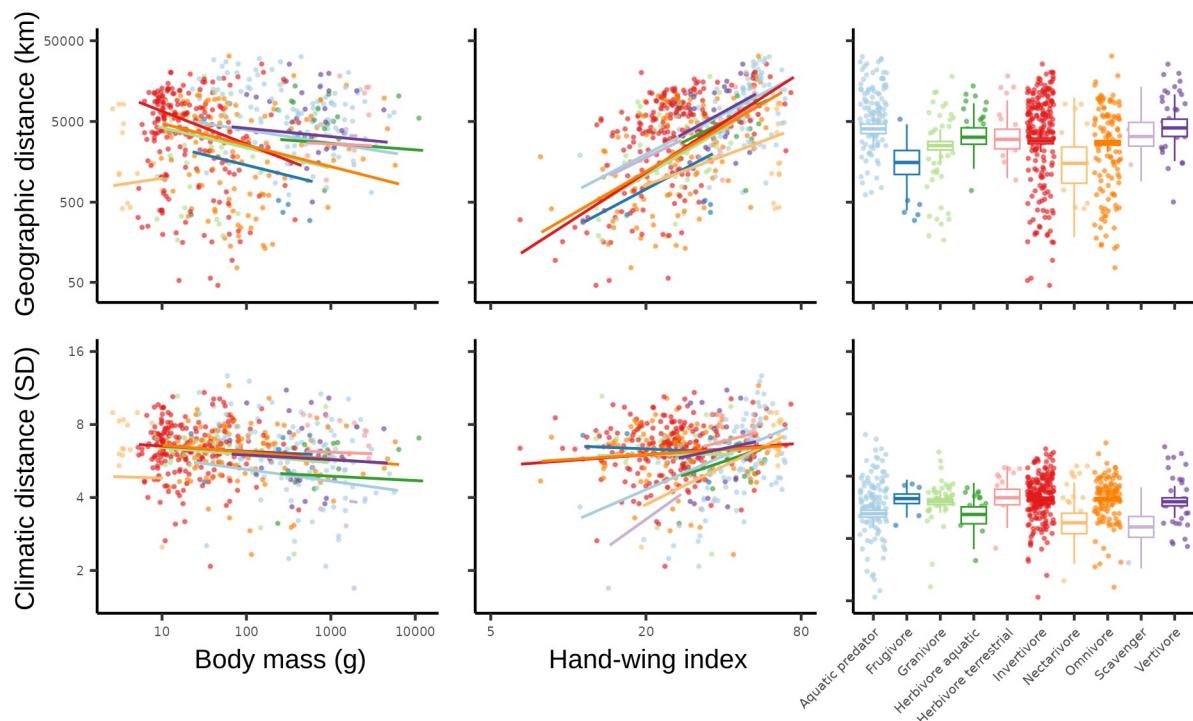
## Table and figures

*Table 4: Effect of each traits on migration distance (A) and climatic distance (B). Univariate regression columns show the total effect of each traits (not controlling for other traits), whereas multivariate regression columns show the direct effect of each traits (controlling for other traits). The values are the mean coefficients and the 95% credible interval from the posterior predictive distribution of the models.*

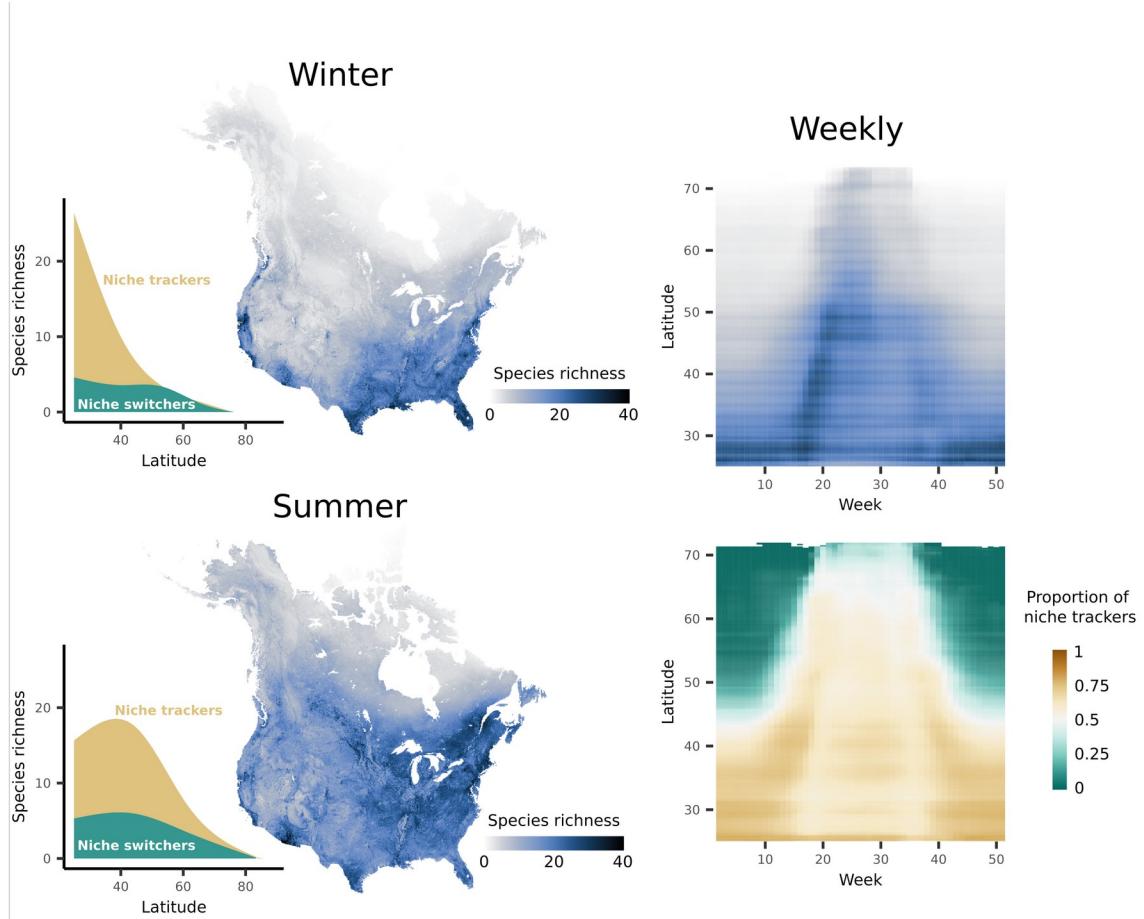
(A)	Migration distance			
	Univariate regression	Multivariate regression		
		Effect	Intercept	Body mass
<b>Trophic guild</b>				
Aquatic predator	8.68 [8.46, 8.91]	8.31 [7.93, 8.73]	-0.30 [-0.57, -0.02]	0.75 [0.49, 0.96]
Frugivore	7.00 [6.05, 7.82]	7.33 [6.40, 8.07]	-0.47 [-1.30, 0.27]	0.76 [0.26, 1.14]
Granivore	7.62 [7.27, 7.99]	7.82 [7.45, 8.15]	-0.46 [-0.94, 0.02]	0.86 [0.57, 1.19]
Herbivore aquatic	8.49 [7.95, 9.04]	8.10 [7.33, 8.90]	-0.15 [-0.66, 0.44]	0.71 [0.22, 1.07]
Herbivore terrestrial	8.35 [7.56, 8.98]	7.99 [6.92, 8.89]	-0.08 [-0.70, 0.66]	0.80 [0.30, 1.38]
Invertivore	8.05 [7.89, 8.21]	7.96 [7.72, 8.19]	-0.74 [-1.03, -0.45]	0.93 [0.77, 1.10]
Nectarivore	7.56 [6.99, 8.13]	7.23 [5.65, 8.40]	0.29 [-0.60, 1.16]	0.54 [-0.07, 0.97]
Omnivore	7.65 [7.43, 7.86]	7.90 [7.70, 8.09]	-0.48 [-0.74, -0.23]	0.83 [0.64, 1.03]
Scavenger	7.90 [6.83, 8.99]	8.17 [6.90, 9.65]	-0.03 [-0.71, 0.84]	0.71 [0.14, 1.15]
Vertivore	8.51 [8.10, 8.91]	8.35 [7.76, 8.99]	-0.19 [-0.73, 0.37]	0.79 [0.36, 1.22]
Body mass	0.12 [0.01, 0.22]			
Hand-wing index	0.63 [0.54, 0.72]			
<b>(B)</b>				
Climatic distance				
(B)	Univariate regression	Multivariate regression		
		Effect	Intercept	Body mass
<b>Trophic guild</b>				
Aquatic predator	1.77 [1.71, 1.82]	1.66 [1.56, 1.77]	-0.09 [-0.16, -0.02]	0.21 [0.13, 0.28]
Frugivore	1.81 [1.64, 1.99]	1.83 [1.66, 2.00]	-0.02 [-0.15, 0.13]	-0.02 [-0.18, 0.13]
Granivore	1.81 [1.73, 1.89]	1.80 [1.70, 1.89]	-0.06 [-0.17, 0.04]	0.01 [-0.10, 0.10]
Herbivore aquatic	1.74 [1.60, 1.86]	1.63 [1.32, 1.85]	-0.03 [-0.13, 0.11]	0.16 [-0.06, 0.41]
Herbivore terrestrial	1.90 [1.77, 2.07]	1.85 [1.59, 2.14]	-0.02 [-0.15, 0.12]	0.13 [-0.15, 0.45]
Invertivore	1.84 [1.80, 1.88]	1.83 [1.77, 1.88]	-0.04 [-0.10, 0.02]	0.04 [-0.01, 0.08]
Nectarivore	1.86 [1.74, 2.00]	1.55 [1.21, 1.82]	-0.02 [-0.13, 0.16]	0.24 [0.06, 0.43]
Omnivore	1.81 [1.76, 1.86]	1.82 [1.77, 1.87]	-0.05 [-0.11, 0.00]	0.03 [-0.03, 0.08]
Scavenger	1.46 [0.85, 1.83]	1.51 [1.15, 1.81]	-0.10 [-0.34, 0.06]	0.34 [0.11, 0.65]
Vertivore	1.81 [1.72, 1.91]	1.79 [1.65, 1.93]	-0.04 [-0.12, 0.07]	0.10 [-0.07, 0.28]
Body mass	-0.05 [-0.07, -0.02]			
Hand-wing index	0.03 [0.00, 0.05]			



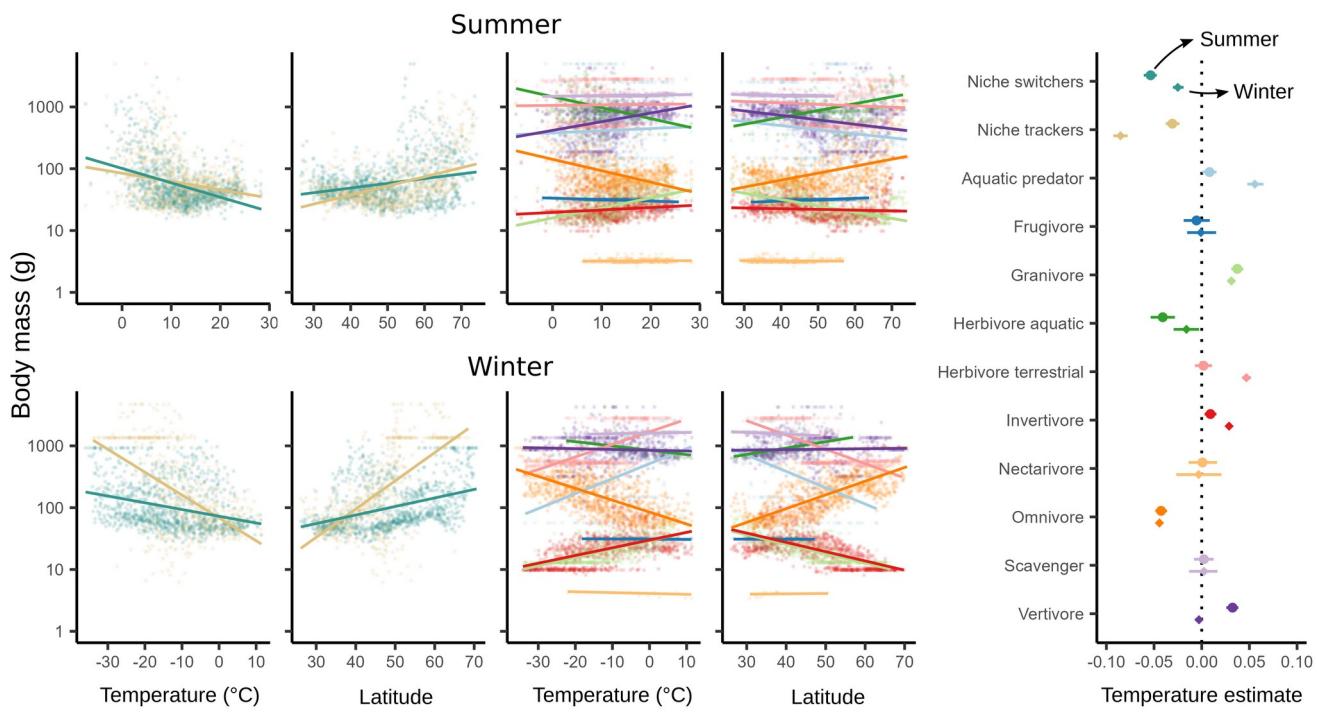
**Figure 10: Trade-off between movements in geographic and climatic space.** Relationship between climatic and geographic distance travelled by each species. Each point is the sum of distances between the centroids of consecutive weeks in climatic and geographic space weighted by abundance. The line is the mean effects while the envelop is the 95% credible interval built using the posterior predictive distribution of the piecewise model. We highlighted an example of resident ‘niche-switcher’ (Rufous-winged Sparrow), resident ‘niche-tracker’ (Great Kiskadee), migrant ‘niche-switcher’ (Buff-breasted Sandpiper), and a migrant ‘niche-tracker’ (Eastern Wood-Pewee). Turquoise points are the highlighted ‘niche-switchers’, and golden points the highlighted ‘niche-trackers’.



**Figure 11: Relationship between migratory behaviour and species traits.** Mean effect of body mass (left panels), hand-wing index (centre panels), and trophic guilds (right panels) on geographic distance (upper panels) and climatic distance (lower panels) travel by birds. Each points is a species, and the lines and boxplot are the effects determined by the posterior predictive distribution. For each trait the figure show the direct effect of the trait (after controlling for other traits).



**Figure 12: Seasonal variation in the latitudinal diversity gradient.** Maps show the species richness calculated as the sum of the probability of occurrences for winter and summer. The area plots show the latitudinal diversity gradient for ‘niche-trackers’ and ‘niche-switchers’. ‘Niche-trackers’ and ‘switchers’ are respectively defined as species that travel less and more than average in climatic space. The upper heatmap shows the average species richness for each week and latitude, whereas the lower heatmap shows the average proportion of niche trackers, calculated as the number of ‘niche-tracker’ species divided by total richness, for each latitude and week.



**Figure 13: Seasonal variation in Bergmann's rule.** Relationship between community weighted mean body mass and temperature and latitude. Models allowed varying slope for tracking strategy ('niche-switchers' in turquoise and 'trackers' in gold) in the two leftmost scatterplot columns or for trophic guilds in the two rightmost scatterplot columns. The upper panels show the mean effects for summer and the lower panels the mean effects for winter. The rightmost plot show the mean effect of temperature on log-transformed body mass with the 95% credible interval built from the posterior distribution.

## Bridging chapter 3 and chapter 4

In chapter 3, I explored seasonal bird biogeography by investigating the influence of traits on bird movement in geographic and within climatic space. As birds migrate, they are not only exposed to different climatic conditions, but also different levels of threats and protections. However, how this variation in threats and protection across seasons explains the decline among bird species has never been explored. Until recently, most studies and protection measures focused on a single season, usually during breeding. Yet, we know that mortality can be at higher rates during nonbreeding and migration, and can have large carryover effects on populations.

In chapter 4, I ask how traits, seasonal level of threats and protection explain the decline in North American birds. I also explore the seasonal and spatial structures of the decline. To do so, I use species-level traits, weekly bird distribution, cumulative human impact, and data on protected areas across North and South America.

# **Chapter 4: The level of human impact and lack of protection during migration drive the decline of the North American avifauna**

**Dominique Caron<sup>1,2</sup>, Pierre Legagneux<sup>3</sup>, David Beauchesne<sup>4</sup>, Phillippe Archambault<sup>3</sup>, Laura J. Pollock<sup>1,2</sup>**

## **Affiliations**

<sup>1</sup> Department of Biology, McGill University, Montreal, QC, Canada

<sup>2</sup> Quebec Centre for Biodiversity Sciences, Montreal, QC, Canada

<sup>3</sup> Département de Biologie and Centre d'études Nordiques, Université Laval, Québec, QC, Canada

<sup>4</sup> University of Toronto, Toronto, Ontario, Canada

The following is in preparation for submission to *Current Biology*

## **Summary**

The decline of North American birds poses a significant threat to ecosystems. Although human pressure is largely responsible for this decline, its impact on species abundance trends throughout the entire annual cycle of birds remains poorly understood. This paper investigates how human footprint and coverage in protected areas throughout the year relate to population trends for 487 North American bird species. We found a decline of 19% and 39% in bird abundances within breeding areas over the entire study area between 1970 and 2017. The only positive trends were seen for the tundra during the wintering season, driven by the increase in abundance of the few overwintering residents. As expected, birds facing more human pressure during migration had more pronounced population declines; this supports evidence suggesting that migration is a critical season for bird conservation and that protecting migration routes and stopovers could play a critical role in mitigating anthropogenic impacts. Contrary to expectations, we observed positive effects of human footprint during breeding or wintering seasons, possibly driven by the increases in urban-tolerant species or ecosystem recoveries of high-human footprint areas. Additionally, we found that smaller birds and those undertaking longer migration distances tend to decline more, highlighting the importance of considering species-specific vulnerabilities in conservation strategies. Overall, our findings underscore the need for holistic conservation approaches that integrate species traits, spatiotemporal variations in stressors and protection measures over birds full annual cycle.

## Introduction

Biodiversity is declining across the globe and is increasingly threatened by environmental pressures driven by human activities (Di Marco et al., 2018; Halpern et al., 2019; Venter et al., 2016; Williams et al., 2020). Therefore, bending the curve of biodiversity loss ultimately depends on our ability to reduce the effect of these human impacts on populations (Leclère et al., 2020). Still, understanding how multiple stressors affect populations is challenging, especially when they vary in time and space, affecting species at single or multiple life stages (Beauchesne et al., 2021; Côté et al., 2016; D. Vinebrooke et al., 2004); this is especially crucial for declining migratory species exposed to varying levels over their annual cycle (Bairlein, 2016; Wilcove & Wikelski, 2008).

Migratory birds are one of such declining groups exposed to different stressors throughout the year (Inger et al., 2015; Lees et al., 2022; Rosenberg et al., 2019). Rosenberg et al. (2019) reported a net loss of 29% of bird abundance in North America between 1970 and 2017. Of course, not all bird species are declining. While the abundance of shorebirds have been declining by 37.4% and 27.1% respectively, the abundance waterfowls have been increasing by 56% (Rosenberg et al., 2019). A major source of mortality in birds are human and industrial activities (Calvert et al., 2013; Loss et al., 2015). It is estimated that domestic cats kill billions of birds in the United States and Canada, and hundreds of millions of birds die from collisions with buildings, automobiles, or power lines each year (Loss et al., 2015). Habitat loss, pesticides, and harvesting have also been identified as the main drivers for the severe decline of farmland birds observed in North America (Stanton et al., 2018). For seabirds, another group experiencing an important decline, invasive species (e.g., rats and cats), incidental mortality in fisheries, and overfishing are the main threats for many species (Dias et al., 2019). So,

birds decline is likely driven, at least partly, to direct human impacts within their breeding, wintering, and migrating ranges.

Recent studies have shown how the level of threats and their effects on population trends can vary across seasons (Buchan et al., 2022; Howard et al., 2020; La Sorte et al., 2022; Vickery et al., 2014; Wilson et al., 2018). For example, La Sorte et al. (2022) showed that most nocturnally migrating North American bird populations were exposed to higher level of air pollution during migration and winter, but did not relate this temporal pattern to long-term population trends. For migratory birds breeding in Europe, Buchan et al. (2022) found that the level of anthropogenic threats during wintering season better predicted long-term population trends, even if it tended to be lower than during the breeding season. In contrast, Howard et al. (2020) found that population trends among European migratory birds are more closely associated with changes in climate than changes in land cover on their breeding grounds. They found the opposite on their wintering ground, where changes in land cover were more closely associated with population trends than changes in climate. So, although the levels of stress birds are exposed to vary during their annual cycle, it remains unclear how they affect population trends, especially during migration seasons.

During migration, birds interrupt their travel between breeding and wintering grounds and aggregate in high density in stopover sites to rest and refuel (Schmaljohann et al., 2022). This creates a large migratory network that populations rely on to fulfill their full annual cycle (Moisan et al., 2023; Zhang et al., 2023). It has been shown that mortality during wintering season and migration can often exceed mortality during the breeding season (Rushing et al., 2017; Sillett & Holmes, 2002). Stressors affecting a species in one season can also carry-over to affect the performance of individuals at other seasons (Norris et al., 2004; Reudink et al., 2009). It has therefore been suggested that habitat quality at

stopover sites is especially crucial for survival and reproductive success (Guo et al., 2023; Morrick et al., 2022; Studds et al., 2017). Yet, most research has focused mainly on one season, usually the breeding season (Faaborg et al., 2010; Marra et al., 2015).

One of the main strategies to slow biodiversity decline is the establishment of protected areas (CBD, 2022). While protected areas have been shown to benefit natural populations (Brighton et al., 2024; Brown et al., 2019; Langhammer et al., 2024; Michel et al., 2021; Sanderson et al., 2023; Wauchope et al., 2022), for migratory species, the success of conservation efforts depends on a network of protected areas (Runge et al., 2014; Zhang et al., 2023). Yet, most Important Bird and Biodiversity Areas, which serve as basis for bird site-based protection worldwide, are based on bird's breeding distributions. Most migratory routes are likely not adequately protected. Runge et al. (2015) showed that, globally, just 9% of 1451 migratory birds are sufficiently covered by protected areas across all stages of their annual cycle, in comparison with 45% for nonmigratory birds. Similarly, Lin et al. (2020) found that less than 10% of stopover sites in North America, which contains around 30% of resident bird abundance in spring and fall migratory periods, were protected. Using a different methodology, Guo et al. (2024) found that landbirds in Northeastern United States tend to aggregate in protected areas as stopovers, but that overall, only a third of high density stopovers are protected and more than half of the protected stopovers are subject to extractive uses. Ultimately, birds need a well-distributed network of well-protected areas, but how the protection level over their full annual cycle relates to population trends is still unclear.

Species responses to human impacts and protection also depend on species traits like habitat specialization, diet, and body size drive species responses to disturbances (Matuoka et al., 2020; Neate-Clegg et al., 2023). For example, Stevens et al. (2024) recently found that smaller body size, habitat

and diet specialization, and longer migration distances were all associated with more severe decline among North American birds. So, birds decline likely depends on the combination of their traits, and their exposure to human pressure within their breeding, wintering, and migrating ranges.

In this study, we consider human pressure and protection over full annual cycle of North American birds to investigate the drivers of trends in bird abundances. To do this, we use weekly and seasonal distributions of species abundances for most bird species in the United States and Canada, along with spatial estimate of cumulative human footprint indices, a cumulative measure of human impacts, on land and water, species-level traits, and estimate of species-level trends. We expect wintering ranges of species to be more impacted but also more protected as many North American birds breed in relatively intact and ‘naturally’ protected Northern regions. We test whether increasing human pressure and lower protection during each season leads to stronger declines. Finally, we test whether species body mass, migratory distances, and diet explain variation in trends. We expect these traits to be important given that many large-bodied species recovered from declines after proactive conservation policies during the last decades (Rosenberg et al., 2019), many long-distance migrants are declining among birds and other groups (Bairlein, 2016; Wilcove & Wikelski, 2008), and that more specialized diet and certain guilds, like invertivores, might be more affected by the decline of their resources (Bowler et al., 2019; Spiller & Dettmers, 2019).

## Results

### *Spatiotemporal patterns of abundance and loss*

The distribution of North American birds varies strongly between breeding, wintering, and migration seasons (seasons of each species defined by eBird). We estimated this by combining the seasonal

distribution of 487 bird species estimated by eBird Status and Trends (Fink et al., 2020, 2023) with their population size and change in abundances between 1970 and 2017 estimated by Rosenberg et al., 2019. We assumed that the number of individual lost for a given species in a given area is proportional to its local density (see Methods for details).

Specifically, birds are more concentrated during the wintering season than other seasons (figure 1d-e). During the wintering season, birds are concentrated in southern and coastal United States, with very few individuals in Northern regions (Figure 1e). In contrast, bird abundances is more dispersed during the breeding season, with the Midwest United States, Northeast United States, Pacific Northwest, and areas around the St. Lawrence River being home of the highest density of birds (figure 1c). During migration, bird abundances show an intermediate distribution between the breeding and wintering seasons, with the US midwest and northeast coast having the highest density (Figure 1d).

Overall, bird abundances decreased in nearly all regions and seasons (figure f-h). On average, bird abundances decreased by 29% (with 50% of the pixel between 19% and 39% decrease) in the breeding season, by 31% (with 50% of the pixel between 24% and 39% decrease) during migration season, and by 23% (with 50% of the pixel between 8% and 37% decrease) during the wintering seasons between 1970 and 2017. During the wintering season, we found an overall increase in bird abundance in the Arctic Tundra. This is driven by the increase of the very few species present in the region during winter. For example, Rosenberg et al. (2019) estimated that the abundances of the Common raven (*Corvus corax*) and the Hoary Redpoll (*Acanthis hornemannii*) have increased by around 80% and 66% respectively between 1970 and 2017.

#### *Seasonal patterns of human footprint and protection*

Overall, we find that birds are exposed to more human pressures during winter than summer (Figure 2a and 2c), but that their range is also better protected during winter (Figure 2b and 2d). This result is likely driven by the lower level of protection and human footprint in Northern regions (figure 1a-b), where the vast majority of species are only present during summer. We estimate this by calculating the mean cumulative human impact on land and water over each species weekly range (Figure 1a) weighted by the abundance. For each bird, we also estimated the level protection for each week by calculating the proportion of species ranges that overlap with protected areas (Figure 1b) weighted by abundances.

We also found distinct spatiotemporal distribution of protection and human impact between declining and increasing birds. For migratory aquatic birds (waterbirds, waterfowls and wetland birds), declining species are exposed to less human pressures on land, more human pressure on water, and are better protected than birds for which the abundance increased (Figure 2a-b). Conversely, for migratory land birds, there is no qualitative difference of human impact among birds declining or increasing in abundance (Figure 2c). However, the ranges of migratory land birds increasing in abundance are better protected than the ones declining, especially during summer (Figure 2d).

#### *Effect of human footprint and protection on trends*

As we expected, we found that seasonal human pressures and protection have significant, but sometimes contrasting effect on migratory birds, and no effect on resident birds (Figure 3). We estimated these effects from modelling species trends as a function the human pressures and level of protection during breeding, wintering, and migration seasons. We consider an effect highly significant when the 95% credible interval (95% CrI) does not contain 0; marginally significant when the 80% credible interval (80% CrI) does not contain 0; and insignificant otherwise.

For migratory land birds, abundance declined less with higher human impact, but only during wintering season (Figure 3a). Human footprint had a highly significant positive effect on trends during wintering (posterior mean: 0.0079; 95% CrI: [0.0029, 0.0130]), a marginally significant positive effect during breeding (posterior mean: 0.0079; 95% CrI: [-0.0002, 0.0159]), but a marginally significant negative effect during migration (posterior mean: -0.0094; 95% CrI: [-0.0189, 0.0003]). In other words, species exposed to greater human pressures during their migration tend to decline more, while species with greater exposure during the breeding and wintering seasons tend to decline less. We also found that species that are better protected during migration are declining less (posterior mean: 0.0061; 95% CrI: [-0.0017, 0.0142]).

For water-based migratory birds, human impact on land had a weaker effect on trends compared to land migratory birds (Figure 3c). However, we found a marginally significant positive effect of human impact on land during migration on species trends (posterior mean: 0.0123; 95% CrI: [-0.0034, 0.0279]). Not surprisingly, human footprint on water was more important than human impact on land for water migratory birds. We found a highly significant negative effect of human footprint during migration (posterior mean: -0.0167; 95% CrI: [-0.0327, 0.0000]), and a highly significant positive effect of human footprint during breeding (posterior mean: 0.0179; 95% CrI: [0.0041, 0.030]). Species for which the wintering ranges were better were declining more (posterior mean: -0.0125; 95% CrI: [-0.0202, -0.0051]).

#### *Traits of declining birds*

We found that smaller species, and species migrating longer distances tended to decline more, demonstrated by modelling trends as a function of traits (Figure 4). Overall, species body mass had the

strongest effect (posterior mean: 0.0119; 95%CrI: [0.0044, 0.0195]). Migration distance had also an effect, but negative and not significant (posterior mean: -0.0045; 95%CrI: [-0.0103, 0.0014]). When controlling for migratory distance and body mass, vertivores (posterior mean: 0.0108; 95%CrI: [-0.0028, 0.0237]) and invertivores (posterior mean: 0.0082; 95%CrI: [-0.0015, 0.018]) declined less than other guilds, although not significantly. This effect for invertivore disappeared when differences in body mass and migratory distance among guilds were not taken into account (posterior mean: -0.0012; 95%CrI: [-0.0086, 0.0061]). Finally, aquatic predators declined more than other guilds (posterior mean: -0.0079; 95%CrI: [-0.0136, -0.0021]), but not only before controlling for differences in body mass and migratory distances (-0.0109; 95%CrI: [-0.0359; 0.0134]).

## Discussion

Migratory bird populations, which are declining globally, are exposed to diverse human-driven stressors and level of protection throughout their annual cycles. In this paper, we investigated the spatiotemporal dynamics of stressors and protection across the annual cycle of North American birds and their association with population trends. Despite a large spatial variation in level of protection and human footprint across the continent, we found that, during breeding season, most bird populations declined between 19% and 39% between 1970 and 2017 throughout North America. Additionally, we found that the level of stress and protection during all seasons, especially during migration, seems to be important drivers of bird trends. Finally, we also found that smaller birds, and birds that travel longer distances tend to decline more than average. These results highlight the importance of considering full annual cycles and species-specific traits in conservation efforts.

We found significant spatiotemporal variation in both human footprint intensity and protection levels across different seasons. On average, birds experience higher levels of human pressures and protection during winter months. This result is not surprising given that many birds breed in the relatively intact, but less protected, boreal and arctic ecosystems and winter in the more densely populated lower latitudes. Previous studies reported similar findings. For example, La Sorte et al. (2022) found that light and air pollution is at its lowest during breeding among nocturnally migrating birds. Wilson et al. (2018) also reported that the Canada Warbler (*Cardellina canadensis*) is exposed to higher human footprint within its wintering range compared to its breeding range. Similar to our finding, Lin et al. (2020) found that breeding grounds are usually less protected than winter ranges.

Importantly, our results demonstrate the significant influence of human footprint and protection on bird population trends, particularly during migration. Birds experiencing higher human pressures during migration exhibit more pronounced declines, demonstrating that the quality of migration routes and stopovers are especially essential to mitigate the negative effects of anthropogenic disturbances on migratory bird populations. This corroborates recent results and calls for focusing management efforts on the quality of bird migration routes, and especially on stopovers (Bayly et al., 2018; Marra et al., 2015; Newton, 2006). For example, Studds et al. (2017) found that the reliance of migratory shorebirds of the East Asian-Australasian Flyway on the Yellow Sea tidal mudflats - where more than 50% of the area has been lost over the past 50 years (Murray et al., 2015) - as stopover sites is a leading driver of population declines. It has also been shown that mortality is often at its highest during migration (Rushing et al., 2017), and that quality of stopovers might have large impacts on the survival and reproductive success during breeding (Legagneux et al., 2011; McGowan et al., 2011). During this relatively long period, birds need to refuel and rest in an already stressful situation due to large

movement cost. Therefore, change in habitat quality can have disproportionate effects that could carry over to affect individuals during other seasons.

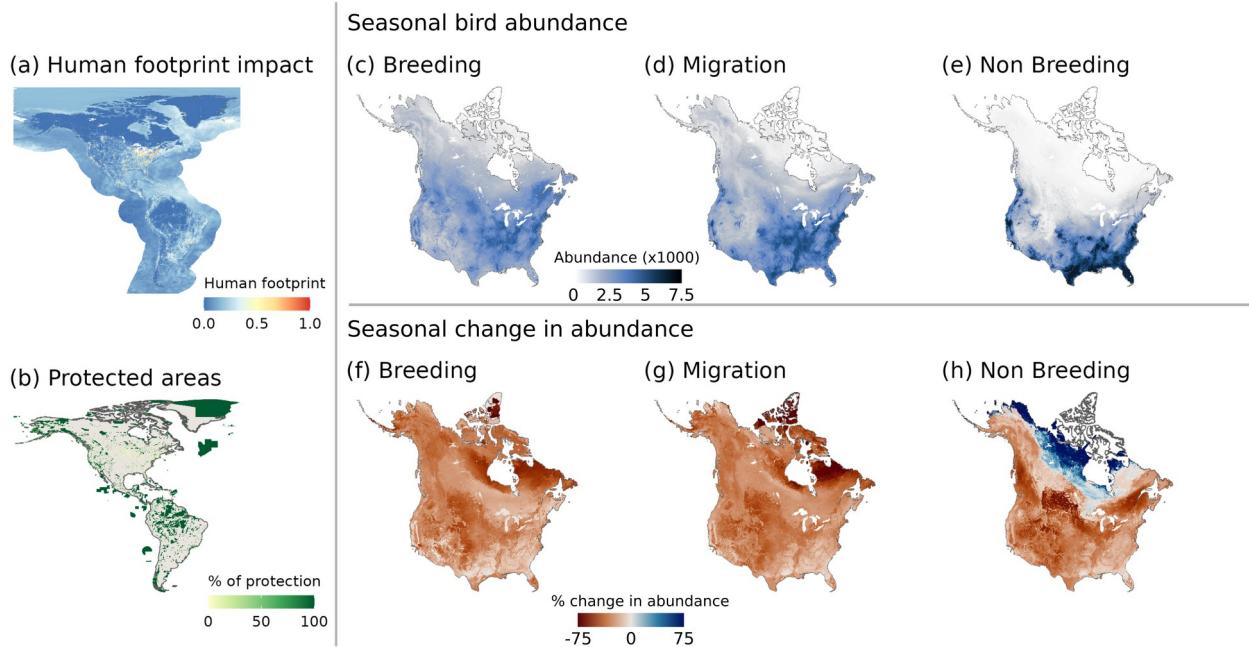
Surprisingly, we found positive effects of human footprint during breeding or nonbreeding season. A few reasons could explain this surprising finding. First, it is possible that urban tolerant species are both declining less and breed/winter more often in high-human footprint environment, driving this relationship. This, however, contrasts with the recent results of Petrenko et al. (2024) who found no relationship between bird urban tolerance and population trend among North American bird species. Second, in North America, areas of high human footprint have been stable or improving in the last decades (Halpern et al., 2019; Venter et al., 2016). So, birds breeding and wintering in high human footprint areas could also be benefiting from recovering ecosystems. Third, correlation between human footprint in different seasons was sometimes very high. It is therefore possible that these positive effect (and the negative effect found for migration) is enhanced by collinearity in our models. However, based on our sensitivity analyses and model diagnostics, our models converged well and coefficients were stable. Finally, Serratosa et al. (2024) recently found that human-induced mortality in the African-Eurasian flyway was highest at intermediate levels of human footprint index. Therefore, it is possible that the relationship between trend and human footprint is non-linear.

Furthermore, our analysis of species traits provides additional insights into the drivers of population trends. Consistent with previous studies (Gilroy et al., 2016; Soykan et al., 2016; Stevens et al., 2024), we found that smaller birds tend to decline more than larger ones. This positive relationship between body mass and species trend contrasts with the general “downsizing” (i.e., loss of larger species) observed for the tetrapods (Atwood et al., 2020; Cooke et al., 2019; Ripple et al., 2017). Soykan et al. (2016) and Stevens et al. (2024) attribute this relationship to the recent recovery of many species, such

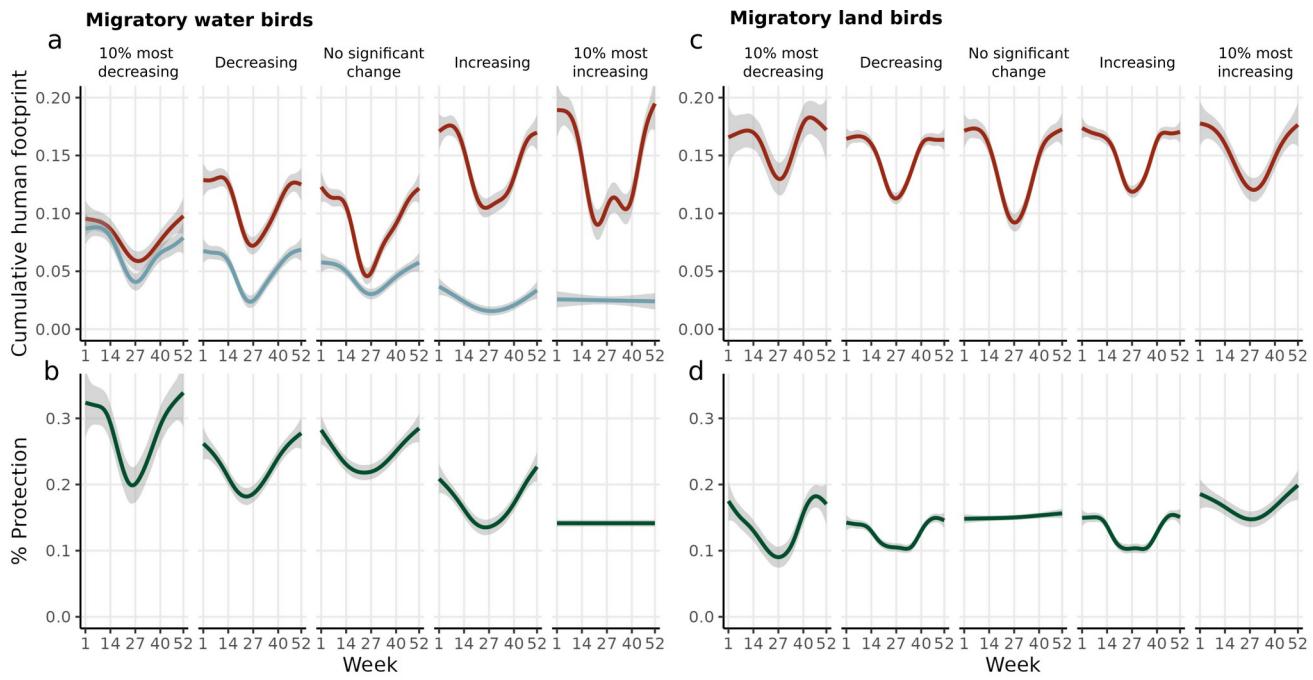
as raptors and waterfowls, following environmental regulations and management efforts. Additionally, birds undertaking longer migration distances also show a trend towards decreased abundance. This is consistent with recent findings that long-distance migrants, which are exposed to additional threats along their migration routes, are declining globally (Horns & Şekercioğlu, 2018; Koleček et al., 2021; Stevens et al., 2024; Wilcove & Wikelski, 2008). Many migratory populations seasonally aggregate in small areas (e.g., staging areas), making them especially vulnerable to perturbations, such as habitat deterioration. They also rely on the connectivity of a large migratory networks, which have been shown to be insufficiently protected (Runge et al., 2015). Finally, climate change may disproportionately affect migrating species by shifting their phenology, migration timing, passage and wintering areas (Visser et al., 2009). This differential responses among trophic guilds, body masses, and migration distance, highlight the loss of functional diversity among birds, having potentially large impact on ecosystem functions.

Overall, our study underscores the importance of adopting a holistic approach to avian conservation that integrates spatiotemporal variation in stressors, protection measures, and species traits. Conservation efforts should prioritize the identification and protection of critical stopover sites and migration corridors, especially in regions experiencing high human pressures. Additionally, targeted conservation interventions tailored to the specific needs and vulnerabilities of different bird species and habitats are essential for effectively mitigating biodiversity loss and promoting the long-term survival of migratory bird populations.

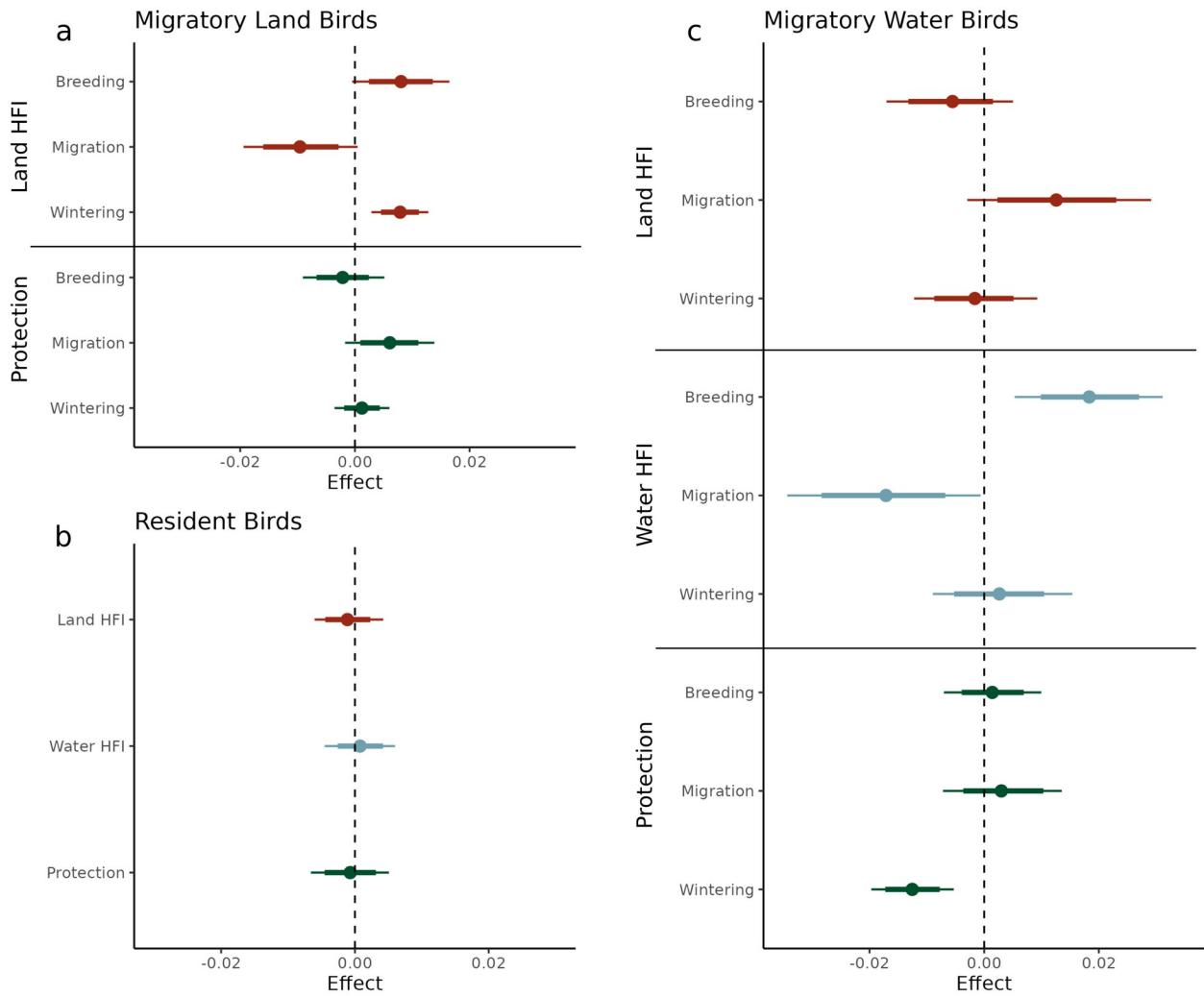
## Figures



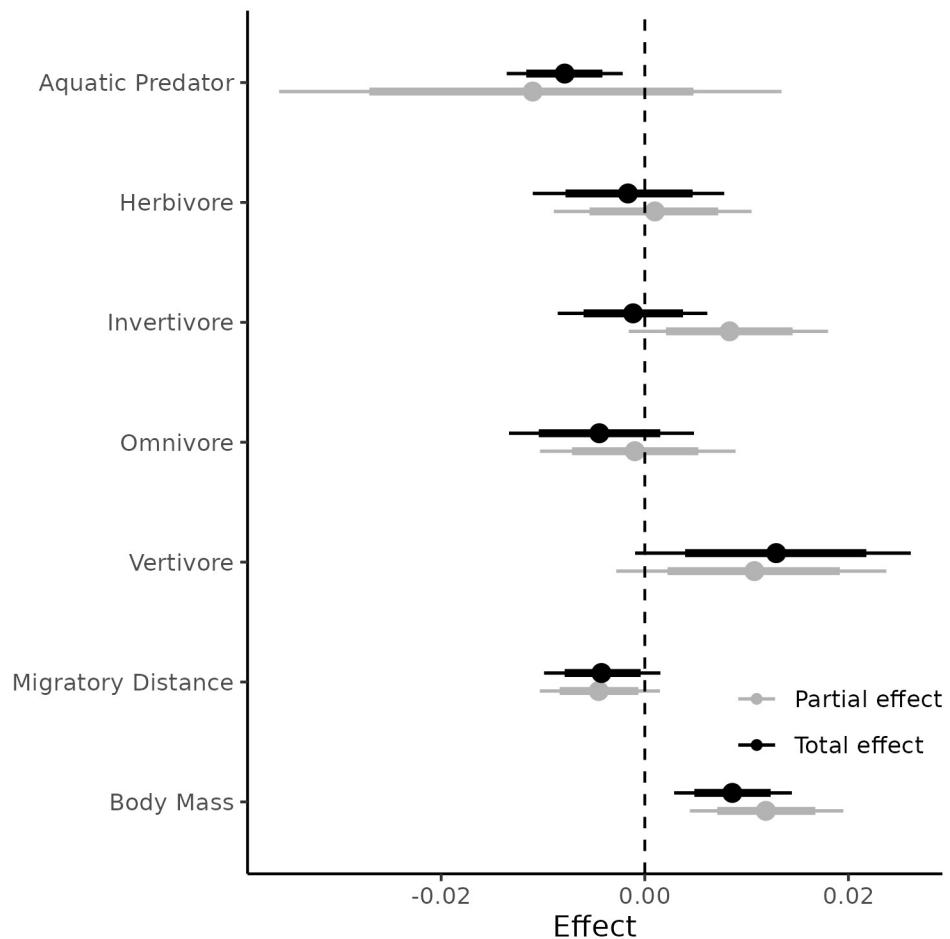
**Figure 14: Seasonal bird abundances and changes in abundances between 1970 and 2017.** a and b, we show the distribution of human footprint and protected areas over North and South America. c-d, we show the number of birds found in each pixel during breeding, migration, and non breeding seasons. f-h show the estimated change in number of individuals in each pixel during breeding, migrating and non breeding seasons when assuming that the local loss of individual of a given species is proportional to its local abundance. For c-f, we used species-specific seasons defined by eBird Status and Trends, and non-migratory birds are included in all three seasons.



**Figure 15: Weekly level of human footprint and protection across migratory water birds and migratory land birds.** In each panel, the line shows the posterior mean and 95% credible interval estimated from generalized additive models; a shows the level of human impact of each week on land (red) and on water (blue) for migratory water birds; b shows the level of protection for each week (green); c shows the level of human impact of each week on land (red) for migratory water birds; and d shows the level of protection for each week (green) for migratory land birds. Species are divided into 5 categories based on their trends between 1970 and 2017.



**Figure 16: Effects of seasonal level of human footprint and protection on species trends.** Point is the posterior mean with their 80% and 95% credible intervals of the effect of each predictor variable on the average yearly proportional change in abundance for species. Predictor variables are log-transformed and scaled.



**Figure 17: Effects of species traits on species trends.** Point is the posterior mean with their 80% and 95% credible intervals of the effect of each predictor variable on the average yearly proportional change in abundance for species. Body mass and migratory distance are log-transformed and scaled.

## **STAR Methods**

### ***Resource availability***

#### *Lead contact*

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Dominique Caron (dominique.caron@mail.mcgill.ca).

#### *Materials availability*

This study did not generate new unique reagents

#### *Data and code availability*

This paper analyzes existing, publicly available data. These accession numbers for the datasets are listed in the key resources table. All original code has been deposited on the Open Science Framework repository ([https://osf.io/emjb3/?view\\_only=d796170b6c304f68b6b545a19b7879df](https://osf.io/emjb3/?view_only=d796170b6c304f68b6b545a19b7879df)) and is publicly available as of the date of publication. DOIs are listed in the key resources table.

## **Method details**

### *Study area*

Most of migratory birds breeding in Canada and the United States of America winters in Southern United States, Mexico, Central America, Caribbean or South America. Consequently, we chose, as a study area, a buffered region around North and South America. We downloaded the World Continents basemap layer from ESRI Data and Maps. We kept the polygons for North America and South America and applied a buffer of 1000 km.

For exploring the spatial variation of bird abundance and loss across seasons, we further constrained our study area to continental United States and Canada. We downloaded Natural Earth v5.0.0's country polygons (<https://www.naturalearthdata.com/downloads/10m-cultural-vectors/10m-admin-0-countries/>), and kept the Canada's and United States of America's polygons. We removed Hawaii and other insular areas in the Caribbean and the Pacific.

### *Bird trends, distributions, and traits*

We focused our analysis on birds breeding in the United States and Canada. We extracted the bird population estimates and trends of 529 species from Rosenberg et al. (2019). Population estimates represent the estimated number of individuals of each species in North America usually between 2006 and 2015. We also extracted the median population trends, which represent the estimated change in the number of individuals for each species usually between 1970 and 2017. Finally, we extracted from Rosenberg et al. (2019) species management group: landbirds, shorebirds, waterbirds, and waterfowl. This gives us one estimate of population size and trends for each species.

Then, for the purpose of understanding how these species-level trends relate to how birds could be experiencing stress across seasons and across space as they move through their full annual cycle, we extracted weekly and seasonal population abundance estimate from eBird Status and Trends data (version 2022) for each of the North American birds using the R package ebirst v3.2022. eBird Status and Trends relates eBird observations to elevation, topography, and land cover in a hierarchical machine learning model that takes into account observation effort and temporal scales to model species distribution for each week. We extracted the relative abundance of each species at the highest resolution available ( $2.8 \times 2.8$  km) within the study area (North and South America). We normalized each layer so that the sum of pixels equals to 1. Values therefore represent the proportion of the

population found in a given pixel at a given time. Models were available for 487 species out of the 529 birds for which we have trend estimates.

For traits, we extracted body mass and diet category for the 487 species from AVONET. AVONET compiles species-level trait means for all the world's bird species. Since some diet categories had very few species, we grouped fugivores, granivores, nectarivores, terrestrial herbivores, and aquatic herbivores into a single diet: herbivores. We also added the only scavenger (*Cathartes aura*) to the vertivores. We also estimated migratory distance of each species by taking the Great Circle Distance of geographical centroids of the breeding and wintering ranges. Geographical centroids of the breeding and wintering ranges was calculated using the mean latitude and longitude weighted by the relative abundance at each season separately.

#### *Human footprint index and protection*

We extracted terrestrial human footprint for 2009 from Venter et al. (2016) and the marine human impact for 2013 from Halpern et al. (2019). In both cases, we extracted values of the cumulative impacts for the study area (buffered North and South America) at a 1km-grid resolution. Cumulative impact combines different stressors into a single estimate of human-driven pressure. Cumulative terrestrial human footprint is a composite of 8 human pressures: built environment, population density, electric infrastructures, crop lands, pasture lands, roads, railways, and navigable waterways. Marine cumulative footprint is a composite of 14 human pressures from fishing (6 layers), climate change (3 layers), shipping (1 layer), pollution (2 layers), light (1 layer), and human presence (1 layer). We standardized both layers of human footprint by dividing values by the respective maximum value within the study areas so that cumulative human impact varies from 0 to 1.

We extracted polygons for protected areas within the Americas from the World Database on Protected Areas (WDPA). We excluded protected areas for which the status is “Proposed” and protected areas for which the year of the status was 2017 or later to match the species trend data. We rasterized and projected the vector layer to the eBird projection, and took the proportion of each pixel within a protected area as value.

### ***Quantification and statistical analysis***

#### *Spatiotemporal abundances and trends*

To understand how bird abundances change across space and seasons, we estimated seasonal abundance of birds for each pixel in North America by taking the product of each species population size estimate and the relative abundance of all birds for each season:

$$\text{Abundance}_{jk} = \sum_{i=1}^{529} \text{RelativeAbundance}_{ijk} \times \text{popest}_i,$$

where  $\text{Abundance}_{jk}$  is the total number of individuals estimated for pixel j during season k,  $\text{RelativeAbundance}_{ijk}$  is the proportion of species i found in pixel j during season k, and  $\text{popest}_i$  is the latest mean population estimate for species i. For each migratory species, we have one abundance estimate for “breeding”, “wintering”, “post-breeding” and “pre-breeding” season, whereas for resident species, we have a single “resident” season.

For migratory species, we summed their abundance over all species for each pixel in each season separately. For resident species, we sum their abundance and then added this layer to each seasonal abundance layer. For migration season, we took the mean between abundance estimate “post-breeding season” and “pre-breeding season”.

To estimate local changes in abundances, we assumed that, for each species, the change in abundance in a given pixel is proportional to its local relative abundance. To do so, we multiplied seasonal relative abundance estimate with each species' trend estimate. We then summed over all species trend layers. Similarly to local abundance layers, we did this for each season separately.

$$AbundanceChange_{jk} = \sum_{i=1}^{529} -RelativeAbundance_{ijk} \times IndividualLost_i,$$

where  $RelativeAbundance_{ijk}$  is the proportion of species  $i$  found in pixel  $j$  during season  $k$ , and  $IndividualsLost$  is the number of individual lost for species  $i$ . The result,  $AbundanceChange_{jk}$ , is an estimate of the absolute number of individuals lost in each pixel and season. We additionally divided the  $AbundanceChange$  layers by the correspondent  $Abundance$  layers to get an estimate of the percentage change in bird abundance for each pixel and season. While ideally, we would have population-level estimates of whether birds are declining in each season and each grid cell, our approach provides a useful approximation by combining species-level trends with level of human impact and protection across space and time.

#### *Seasonality of protection and human impacts*

For each species, we estimated the exposition to weekly cumulative impact on terrestrial and marine areas for each week and season. To do so, we assume that the level of human footprint within a pixel is constant within a year. We multiplied the relative abundance estimate of each bird and time with the human footprint layer. We then summed over the values of each layer. This results in the average human footprint experience by an individual of a given species at a given time.

We also estimated the level of protection for each species and week and season. To do so, we multiplied the relative abundance estimate of each bird and time with the protected area layer, and then summed

over the values of each layer. This results in an estimate of the percentage of individuals found in a protected area for a given species at a given time.

To explore the level of human footprint and protection experienced by birds in each season, we fitted a generalized additive model with week as predictor. We did this analysis for water migratory birds (waterfowls, waterbirds, and shorebirds) and land migratory birds (landbirds) separately. For water migratory birds, we separated human footprint on terrestrial and marine areas since they were extracted from two different sources and assembled using different methodologies. For land migratory birds, we excluded human footprint on marine areas since none of their range was found on water for most species. We excluded resident birds from the temporal analysis since the level of human footprint and protection do not vary over time. We further separated species into 5 subcategories according to their trends to estimate if temporal dynamics of human footprint and protection qualitatively differ between increasing and declining birds. We first identified the 10% most declining and 10% most increasing birds in terms of percentage change in abundance between 1970 and 2017, and used the 95% credible interval reported by Rosenberg et al. (2019) to identify species that are significantly increasing and declining.

#### *Trends-impact and protection relationships*

For each species, we standardized trends by dividing the change in abundance by the number of years the change documents and by the population estimate for the species:

$$Trend_i = -\frac{IndividualsLost_i}{(LastYear - FirstYear) \times popest_i},$$

where  $IndividualsLost$  is the number of individual lost between  $FirstYear$  and  $LastYear$ , and  $popest$  is the latest mean population estimate for species  $i$ . Therefore,  $Trend_i$  represents the average yearly proportional change in abundance for species  $i$  in North America.

We modelled trend as a function of level of protection and human footprint for land migratory birds, water migratory birds, and resident birds separately. For land migratory birds, we modelled trend as a function of the level of protection and human footprint on land during the breeding, the wintering, and migration season (six fixed effects). For water migratory birds, we added fixed effects for the level of human footprint during breeding, wintering, and migration season (nine fixed effects in total), and a random intercept for bird management group (three levels: waterfowl, waterbird, shorebird). Finally, for resident birds, the model only had three fixed effects: level of protection, land human footprint, and water human footprint. For each models, all fixed effects (level of protection and human footprints) were log-transformed and scaled.

### *Trends-trait relationship*

We also modelled trend as a function species traits. We considered three traits: bird diet, body mass, and migratory distance. Diet was a categorical variable with 5 levels: Aquatic predator (109 species), herbivore (61 species), invertivore (146 species), omnivore (71 species), and vertivore (22 species). Migratory distance and body mass were log-transformed and standardized by dividing by twice their standard deviation to make coefficients of quantitative and categorical predictors comparable (Gelman, 2008).

Parameters of all models were estimated with Hamiltonian Monte Carlo (Neal, 2011). We ran four chains, each with 1000 warm-up iterations, followed by 1000 iterations for inference. We diagnosed

convergence and adequacy with trace rank plots, posterior predictive checks, and we calculated the rank-normalized potential scale reduction factor on split chains for all runs (Vehtari et al., 2021). We conducted the analyses using Stan (Carpenter et al., 2017) through the package brms in R (Bürkner, 2018).

## **Key Resource Table**

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<i>Deposited data</i>		
Bird distributions	Fink et al., 2023	<a href="https://doi.org/10.2173/ebirdst.2022">https://doi.org/10.2173/ebirdst.2022</a>
Bird trends	Rosenberg et al., 2019	<a href="https://doi.org/10.1126/science.aaw1313">https://doi.org/10.1126/science.aaw1313</a>
Bird traits	Tobias et al., 2022	<a href="https://doi.org/10.1111/ele.13898">https://doi.org/10.1111/ele.13898</a>
Land human footprint	Venter et al., 2017	<a href="https://doi.org/10.5061/dryad.052q5">https://doi.org/10.5061/dryad.052q5</a>
Marine human footprint	Frazier, 2019	<a href="https://doi.org/10.5063/F12B8WBS">https://doi.org/10.5063/F12B8WBS</a>
Protected areas	UNEP-WCMC & IUCN (2023)	N/A
Country boundaries version 5.0.0	Natural Earth	N/A
Continent boundaries	Esri Data and Maps	N/A
Analyzed data	This paper	<a href="https://osf.io/emjb3/?view_only=d796170b6c304f68b6b545a19b7879df">https://osf.io/emjb3/?view_only=d796170b6c304f68b6b545a19b7879df</a>
<i>Software and algorithms</i>		
R Statistical Computing language version 4.3.3	R Core Team, 2024	<a href="https://cran.r-project.org/src/base/R-4/R-4.3.3.tar.gz">https://cran.r-project.org/src/base/R-4/R-4.3.3.tar.gz</a>
ebirdst R package version 3.2022	Strimas-Mackey et al., 2023	<a href="https://ebird.github.io/ebirdst/">https://ebird.github.io/ebirdst/</a>
terra R package version 1.7-55	Hijmans, 2023	<a href="https://CRAN.R-project.org/package=terra">https://CRAN.R-project.org/package=terra</a>
brms R package 2.20.6	Bürkner, 2017, 2018	<a href="https://doi.org/10.18637/jss.v080.i01">https://doi.org/10.18637/jss.v080.i01</a>
Code to replicate analyses of this study	This paper	<a href="https://osf.io/emjb3/?view_only=d796170b6c304f68b6b545a19b7879df">https://osf.io/emjb3/?view_only=d796170b6c304f68b6b545a19b7879df</a>

## References

- Atwood, T. B., Valentine, S. A., Hammill, E., McCauley, D. J., Madin, E. M. P., Beard, K. H., & Pearse, W. D. (2020). Herbivores at the highest risk of extinction among mammals, birds, and reptiles. *Science Advances*, 6(32), eabb8458. <https://doi.org/10.1126/sciadv.abb8458>
- Bairlein, F. (2016). Migratory birds under threat. *Science*, 354(6312), 547–548. <https://doi.org/10.1126/science.aah6647>
- Bayly, N. J., Rosenberg, K. V., Easton, W. E., Gómez, C., Carlisle, J., Ewert, D. N., Drake, A., & Goodrich, L. (2018). Major stopover regions and migratory bottlenecks for Nearctic-Neotropical landbirds within the Neotropics: A review. *Bird Conservation International*, 28(1), 1–26. <https://doi.org/10.1017/S0959270917000296>
- Beauchesne, D., Cazelles, K., Archambault, P., Dee, L. E., & Gravel, D. (2021). On the sensitivity of food webs to multiple stressors. *Ecology Letters*, 24(10), 2219–2237. <https://doi.org/10.1111/ele.13841>
- Bowler, D. E., Heldbjerg, H., Fox, A. D., de Jong, M., & Böhning-Gaese, K. (2019). Long-term declines of European insectivorous bird populations and potential causes. *Conservation Biology*, 33(5), 1120–1130. <https://doi.org/10.1111/cobi.13307>
- Brighton, C. H., Massimino, D., Boersch-Supan, P., Barnes, A. E., Martay, B., Bowler, D. E., Hoskins, H. M. J., & Pearce-Higgins, J. W. (2024). The benefits of protected areas for bird population trends may depend on their condition. *Biological Conservation*, 292, 110553. <https://doi.org/10.1016/j.biocon.2024.110553>
- Brown, J. A., Lockwood, J. L., Avery, J. D., Curtis Burkhalter, J., Aagaard, K., & Fenn, K. H. (2019). Evaluating the long-term effectiveness of terrestrial protected areas: A 40-year look at forest

- bird diversity. *Biodiversity and Conservation*, 28(4), 811–826. <https://doi.org/10.1007/s10531-018-01693-5>
- Buchan, C., Franco, A. M. A., Catry, I., Gamero, A., Klvaňová, A., & Gilroy, J. J. (2022). Spatially explicit risk mapping reveals direct anthropogenic impacts on migratory birds. *Global Ecology and Biogeography*, 31(9), 1707–1725. <https://doi.org/10.1111/geb.13551>
- Bürkner, P.-C. (2017). **brms**: An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software*, 80(1). <https://doi.org/10.18637/jss.v080.i01>
- Bürkner, P.-C. (2018). Advanced Bayesian Multilevel Modeling with the R Package brms. *The R Journal*, 10(1), 395. <https://doi.org/10.32614/RJ-2018-017>
- Calvert, A. M., Bishop, C. A., Elliot, R. D., Krebs, E. A., Kydd, T. M., Machtans, C. S., & Robertson, G. J. (2013). A Synthesis of Human-related Avian Mortality in Canada. *Avian Conservation and Ecology*, 8(2), art11. <https://doi.org/10.5751/ACE-00581-080211>
- Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., & Riddell, A. (2017). Stan: A Probabilistic Programming Language. *Journal of Statistical Software*, 76(1). <https://doi.org/10.18637/jss.v076.i01>
- CBD. (2022). *The Kunming-Montreal Global Biodiversity Framework*. Conference of the Parties to the Convention on Biological Diversity, Montreal, Canada.  
<https://www.cbd.int/doc/c/e6d3/cd1d/daf663719a03902a9b116c34/cop-15-l-25-en.pdf>
- Cooke, R. S. C., Eigenbrod, F., & Bates, A. E. (2019). Projected losses of global mammal and bird ecological strategies. *Nature Communications*, 10(1), 2279. <https://doi.org/10.1038/s41467-019-10284-z>
- Côté, I. M., Darling, E. S., & Brown, C. J. (2016). Interactions among ecosystem stressors and their importance in conservation. *Proceedings of the Royal Society B: Biological Sciences*, 283(1824), 20152592. <https://doi.org/10.1098/rspb.2015.2592>

- D. Vinebrooke, R., L. Cottingham, K., Norberg, J., Marten Scheffer, I. Dodson, S., C. Maberly, S., & Sommer, U. (2004). Impacts of multiple stressors on biodiversity and ecosystem functioning: The role of species co-tolerance. *Oikos*, 104(3), 451–457. <https://doi.org/10.1111/j.0030-1299.2004.13255.x>
- Di Marco, M., Venter, O., Possingham, H. P., & Watson, J. E. M. (2018). Changes in human footprint drive changes in species extinction risk. *Nature Communications*, 9(1), 4621. <https://doi.org/10.1038/s41467-018-07049-5>
- Dias, M. P., Martin, R., Pearmain, E. J., Burfield, I. J., Small, C., Phillips, R. A., Yates, O., Lascelles, B., Borboroglu, P. G., & Croxall, J. P. (2019). Threats to seabirds: A global assessment. *Biological Conservation*, 237, 525–537. <https://doi.org/10.1016/j.biocon.2019.06.033>
- Faaborg, J., Holmes, R. T., Anders, A. D., Bildstein, K. L., Dugger, K. M., Gauthreaux Jr., S. A., Heglund, P., Hobson, K. A., Jahn, A. E., Johnson, D. H., Latta, S. C., Levey, D. J., Marra, P. P., Merkord, C. L., Nol, E., Rothstein, S. I., Sherry, T. W., Sillett, T. S., Thompson III, F. R., & Warnock, N. (2010). Conserving migratory land birds in the New World: Do we know enough? *Ecological Applications*, 20(2), 398–418. <https://doi.org/10.1890/09-0397.1>
- Fink, D., Auer, T., Johnston, A., Ruiz-Gutierrez, V., Hochachka, W. M., & Kelling, S. (2020). Modeling avian full annual cycle distribution and population trends with citizen science data. *Ecological Applications*, 30(3). <https://doi.org/10.1002/eap.2056>
- Fink, D., Auer, T., Johnston, A., Strimas-Mackey, M., Ligocki, S., Robinson, O., Hochachka, W., Jaromczyk, L., Crowley, C., Dunham, K., Stillman, A., Davies, I., Rodewald, A., Ruiz-Gutierrez, V., & Wood, C. (2023). *eBird Status and Trends* [dataset]. <https://doi.org/10.2173/ebirdst.2022>
- Frazier, M. (2019). *Recent pace of change in human impact on the world's ocean: Cumulative impacts* [Text/xml]. [object Object]. <https://doi.org/10.5063/F12B8WBS>

- Gelman, A. (2008). Scaling regression inputs by dividing by two standard deviations. *Statistics in Medicine*, 27(15), 2865–2873. <https://doi.org/10.1002/sim.3107>
- Gilroy, J. J., Gill, J. A., Butchart, S. H. M., Jones, V. R., & Franco, A. M. A. (2016). Migratory diversity predicts population declines in birds. *Ecology Letters*, 19(3), 308–317. <https://doi.org/10.1111/ele.12569>
- Guo, F., Buler, J. J., Smolinsky, J. A., & Wilcove, D. S. (2023). Autumn stopover hotspots and multiscale habitat associations of migratory landbirds in the eastern United States. *Proceedings of the National Academy of Sciences*, 120(3), e2203511120. <https://doi.org/10.1073/pnas.2203511120>
- Guo, F., Buler, J. J., Smolinsky, J. A., & Wilcove, D. S. (2024). Seasonal patterns and protection status of stopover hotspots for migratory landbirds in the eastern United States. *Current Biology*, 34(2), 235-244.e3. <https://doi.org/10.1016/j.cub.2023.11.033>
- Halpern, B. S., Frazier, M., Afflerbach, J., Lowndes, J. S., Micheli, F., O’Hara, C., Scarborough, C., & Selkoe, K. A. (2019). Recent pace of change in human impact on the world’s ocean. *Scientific Reports*, 9(1), 11609. <https://doi.org/10.1038/s41598-019-47201-9>
- Hijmans, R. J. (2023). *terra: Spatial Data Analysis*. <https://CRAN.R-project.org/package=terra>
- Horns, J. J., & Şekercioğlu, Ç. H. (2018). Conservation of migratory species. *Current Biology*, 28(17), R980–R983. <https://doi.org/10.1016/j.cub.2018.06.032>
- Howard, C., Stephens, P. A., Pearce-Higgins, J. W., Gregory, R. D., Butchart, S. H. M., & Willis, S. G. (2020). Disentangling the relative roles of climate and land cover change in driving the long-term population trends of European migratory birds. *Diversity and Distributions*, 26(11), 1442–1455. <https://doi.org/10.1111/ddi.13144>
- Inger, R., Gregory, R., Duffy, J. P., Stott, I., Voršík, P., & Gaston, K. J. (2015). Common European birds are declining rapidly while less abundant species’ numbers are rising. *Ecology Letters*,

18(1), 28–36. <https://doi.org/10.1111/ele.12387>

Koleček, J., Reif, J., Šálek, M., Hanzelka, J., Sottas, C., & Kubelka, V. (2021). Global population trends in shorebirds: Migratory behaviour makes species at risk. *The Science of Nature*, 108(2), 9. <https://doi.org/10.1007/s00114-021-01717-1>

La Sorte, F. A., Horton, K. G., Johnston, A., Fink, D., & Auer, T. (2022). Seasonal associations with light pollution trends for nocturnally migrating bird populations. *Ecosphere*, 13(3). <https://doi.org/10.1002/ecs2.3994>

Langhammer, P. F., Bull, J. W., Bicknell, J. E., Oakley, J. L., Brown, M. H., Bruford, M. W., Butchart, S. H. M., Carr, J. A., Church, D., Cooney, R., Cutajar, S., Foden, W., Foster, M. N., Gascon, C., Geldmann, J., Genovesi, P., Hoffmann, M., Howard-McCombe, J., Lewis, T., ... Brooks, T. M. (2024). The positive impact of conservation action. *Science*, 384(6694), 453–458.

<https://doi.org/10.1126/science.adj6598>

Leclère, D., Obersteiner, M., Barrett, M., Butchart, S. H. M., Chaudhary, A., De Palma, A., DeClerck, F. A. J., Di Marco, M., Doelman, J. C., Dürauer, M., Freeman, R., Harfoot, M., Hasegawa, T., Hellweg, S., Hilbers, J. P., Hill, S. L. L., Humpenöder, F., Jennings, N., Krisztin, T., ... Young, L. (2020). Bending the curve of terrestrial biodiversity needs an integrated strategy. *Nature*, 585(7826), 551–556. <https://doi.org/10.1038/s41586-020-2705-y>

Lees, A. C., Haskell, L., Allinson, T., Bezeng, S. B., Burfield, I. J., Renjifo, L. M., Rosenberg, K. V., Viswanathan, A., & Butchart, S. H. M. (2022). State of the World's Birds. *Annual Review of Environment and Resources*, 47(Volume 47, 2022), 231–260. <https://doi.org/10.1146/annurev-environ-112420-014642>

Legagneux, P., Fast, P. L. F., Gauthier, G., & Béty, J. (2011). Manipulating individual state during migration provides evidence for carry-over effects modulated by environmental conditions.

*Proceedings of the Royal Society B: Biological Sciences*, 279(1730), 876–883.

<https://doi.org/10.1098/rspb.2011.1351>

Lin, H.-Y., Schuster, R., Wilson, S., Cooke, S. J., Rodewald, A. D., & Bennett, J. R. (2020). Integrating season-specific needs of migratory and resident birds in conservation planning. *Biological Conservation*, 252, 108826. <https://doi.org/10.1016/j.biocon.2020.108826>

Loss, S. R., Will, T., & Marra, P. P. (2015). Direct Mortality of Birds from Anthropogenic Causes. *Annual Review of Ecology, Evolution and Systematics*, 46(Volume 46, 2015), 99–120.

<https://doi.org/10.1146/annurev-ecolsys-112414-054133>

Marra, P. P., Cohen, E. B., Loss, S. R., Rutter, J. E., & Tonra, C. M. (2015). A call for full annual cycle research in animal ecology. *Biology Letters*, 11(8), 20150552.

<https://doi.org/10.1098/rsbl.2015.0552>

Matuoka, M. A., Benchimol, M., Almeida-Rocha, J. M. de, & Morante-Filho, J. C. (2020). Effects of anthropogenic disturbances on bird functional diversity: A global *meta*-analysis. *Ecological Indicators*, 116, 106471. <https://doi.org/10.1016/j.ecolind.2020.106471>

McGowan, C. P., Hines, J. E., Nichols, J. D., Lyons, J. E., Smith, D. R., Kalasz, K. S., Niles, L. J., Dey, A. D., Clark, N. A., Atkinson, P. W., Minton, C. D. T., & Kendall, W. (2011). Demographic consequences of migratory stopover: Linking red knot survival to horseshoe crab spawning abundance. *Ecosphere*, 2(6), art69. <https://doi.org/10.1890/ES11-00106.1>

Michel, N. L., Saunders, S. P., Meehan, T. D., & Wilsey, C. B. (2021). Effects of stewardship on protected area effectiveness for coastal birds. *Conservation Biology*, 35(5), 1484–1495.

<https://doi.org/10.1111/cobi.13698>

Moisan, L., Gravel, D., Legagneux, P., Gauthier, G., Léandri-Breton, D.-J., Somveille, M., Therrien, J.-F., Lamarre, J.-F., & Béty, J. (2023). Scaling migrations to communities: An empirical case of

- migration network in the Arctic. *Frontiers in Ecology and Evolution*, 10.  
<https://www.frontiersin.org/articles/10.3389/fevo.2022.1077260>
- Morriss, Z. N., Lilleyman, A., Fuller, R. A., Bush, R., Coleman, J. T., Garnett, S. T., Gerasimov, Y. N., Jessop, R., Ma, Z., Maglio, G., Minton, C. D. T., Syroechkovskiy, E., & Woodworth, B. K. (2022). Differential population trends align with migratory connectivity in an endangered shorebird. *Conservation Science and Practice*, 4(1), e594. <https://doi.org/10.1111/csp2.594>
- Murray, N. J., Ma, Z., & Fuller, R. A. (2015). Tidal flats of the Yellow Sea: A review of ecosystem status and anthropogenic threats. *Austral Ecology*, 40(4), 472–481.  
<https://doi.org/10.1111/aec.12211>
- Neal, R. M. (2011). MCMC Using Hamiltonian Dynamics. In *Handbook of Markov Chain Monte Carlo* (pp. 113–162). CRC Press. <https://doi.org/10.1201/b10905-7>
- Neate-Clegg, M. H. C., Tonelli, B. A., Youngflesh, C., Wu, J. X., Montgomery, G. A., Şekercioğlu, Ç. H., & Tingley, M. W. (2023). Traits shaping urban tolerance in birds differ around the world. *Current Biology*, 33(9), 1677-1688.e6. <https://doi.org/10.1016/j.cub.2023.03.024>
- Newton, I. (2006). Can conditions experienced during migration limit the population levels of birds? *Journal of Ornithology*, 147(2), 146–166. <https://doi.org/10.1007/s10336-006-0058-4>
- Norris, D. R., Marra, P. P., Kyser, T. K., Sherry, T. W., & Ratcliffe, L. M. (2004). Tropical winter habitat limits reproductive success on the temperate breeding grounds in a migratory bird. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 271(1534), 59–64.  
<https://doi.org/10.1098/rspb.2003.2569>
- Petrenko, J. A., Martin, P. R., Fanelli, R. E., & Bonier, F. (2024). Urban tolerance does not protect against population decline in North American birds. *Biology Letters*, 20(1), 20230507.  
<https://doi.org/10.1098/rsbl.2023.0507>

- R Core Team. (2024). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Reudink, M. W., Marra, P. P., Kyser, T. K., Boag, P. T., Langin, K. M., & Ratcliffe, L. M. (2009). Non-breeding season events influence sexual selection in a long-distance migratory bird. *Proceedings of the Royal Society B: Biological Sciences*, 276(1662), 1619–1626. <https://doi.org/10.1098/rspb.2008.1452>
- Ripple, W. J., Wolf, C., Newsome, T. M., Hoffmann, M., Wirsing, A. J., & McCauley, D. J. (2017). Extinction risk is most acute for the world's largest and smallest vertebrates. *Proceedings of the National Academy of Sciences*, 114(40), 10678–10683. <https://doi.org/10.1073/pnas.1702078114>
- Rosenberg, K. V., Dokter, A. M., Blancher, P. J., Sauer, J. R., Smith, A. C., Smith, P. A., Stanton, J. C., Panjabi, A., Helft, L., Parr, M., & Marra, P. P. (2019). Decline of the North American avifauna. *Science*, 366(6461), 120–124. <https://doi.org/10.1126/science.aaw1313>
- Runge, C. A., Martin, T. G., Possingham, H. P., Willis, S. G., & Fuller, R. A. (2014). Conserving mobile species. *Frontiers in Ecology and the Environment*, 12(7), 395–402. <https://doi.org/10.1890/130237>
- Runge, C. A., Watson, J. E. M., Butchart, S. H. M., Hanson, J. O., Possingham, H. P., & Fuller, R. A. (2015). Protected areas and global conservation of migratory birds. *Science*, 350(6265), 1255–1258. <https://doi.org/10.1126/science.aac9180>
- Rushing, C. S., Hostetler, J. A., Sillett, T. S., Marra, P. P., Rotenberg, J. A., & Ryder, T. B. (2017). Spatial and temporal drivers of avian population dynamics across the annual cycle. *Ecology*, 98(11), 2837–2850. <https://doi.org/10.1002/ecy.1967>
- Sanderson, F. J., Wilson, J. D., Franks, S. E., & Buchanan, G. M. (2023). Benefits of protected area networks for breeding bird populations and communities. *Animal Conservation*, 26(3), 279–

289. <https://doi.org/10.1111/acv.12832>

Schmaljohann, H., Eikenaar, C., & Sapir, N. (2022). Understanding the ecological and evolutionary function of stopover in migrating birds. *Biological Reviews*, 97(4), 1231–1252.

<https://doi.org/10.1111/brv.12839>

Serratosa, J., Oppel, S., Rotics, S., Santangeli, A., Butchart, S. H. M., Cano-Alonso, L. S., Tellería, J. L., Kemp, R., Nicholas, A., Kalvāns, A., Galarza, A., Franco, A. M. A., Andreotti, A., Kirschel, A. N. G., Ngari, A., Soutullo, A., Bermejo-Bermejo, A., Botha, A. J., Ferri, A., ... Jones, V. R. (2024). Tracking data highlight the importance of human-induced mortality for large migratory birds at a flyway scale. *Biological Conservation*, 293, 110525.

<https://doi.org/10.1016/j.biocon.2024.110525>

Sillett, T. S., & Holmes, R. T. (2002). Variation in survivorship of a migratory songbird throughout its annual cycle. *Journal of Animal Ecology*, 71(2), 296–308. <https://doi.org/10.1046/j.1365-2656.2002.00599.x>

Soykan, C. U., Sauer, J., Schuetz, J. G., LeBaron, G. S., Dale, K., & Langham, G. M. (2016). Population trends for North American winter birds based on hierarchical models. *Ecosphere*, 7(5), e01351. <https://doi.org/10.1002/ecs2.1351>

Spiller, K. J., & Dettmers, R. (2019). Evidence for multiple drivers of aerial insectivore declines in North America. *The Condor*, 121(2), duz010. <https://doi.org/10.1093/condor/duz010>

Stanton, R. L., Morrissey, C. A., & Clark, R. G. (2018). Analysis of trends and agricultural drivers of farmland bird declines in North America: A review. *Agriculture, Ecosystems & Environment*, 254, 244–254. <https://doi.org/10.1016/j.agee.2017.11.028>

Stevens, H. C., Smith, A. C., Buechley, E. R., Şekercioğlu, Ç. H., Shirey, V., Rosenberg, K. V., La Sorte, F. A., Tallamy, D., & Marra, P. P. (2024). Species-specific ecological traits, phylogeny,

and geography underpin vulnerability to population declines for North American birds. *Ornithological Applications*, 126(1), duad046. <https://doi.org/10.1093/ornithapp/duad046>

Strimas-Mackey, M., Ligocki, S., Auer, T., & Fink, D. (2023). *ebirdst: Access and Analyze eBird Status and Trends Data Products*. <https://ebird.github.io/ebirdst/>

Studds, C. E., Kendall, B. E., Murray, N. J., Wilson, H. B., Rogers, D. I., Clemens, R. S., Gosbell, K., Hassell, C. J., Jessop, R., Melville, D. S., Milton, D. A., Minton, C. D. T., Possingham, H. P., Riegen, A. C., Straw, P., Woehler, E. J., & Fuller, R. A. (2017). Rapid population decline in migratory shorebirds relying on Yellow Sea tidal mudflats as stopover sites. *Nature Communications*, 8(1), 14895. <https://doi.org/10.1038/ncomms14895>

Tobias, J. A., Sheard, C., Pigot, A. L., Devenish, A. J. M., Yang, J., Sayol, F., Neate-Clegg, M. H. C., Alioravainen, N., Weeks, T. L., Barber, R. A., Walkden, P. A., MacGregor, H. E. A., Jones, S. E. I., Vincent, C., Phillips, A. G., Marples, N. M., Montaño-Centellas, F. A., Leandro-Silva, V., Claramunt, S., ... Schleuning, M. (2022). AVONET: Morphological, ecological and geographical data for all birds. *Ecology Letters*, 25(3), 581–597. <https://doi.org/10.1111/ele.13898>

Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., & Bürkner, P.-C. (2021). Rank-Normalization, Folding, and Localization: An Improved  $R^*$  for Assessing Convergence of MCMC (with Discussion). *Bayesian Analysis*, 16(2), 667–718. <https://doi.org/10.1214/20-BA1221>

Venter, O., Sanderson, E. W., Magrach, A., Allan, J. R., Beher, J., Jones, K. R., Possingham, H. P., Laurance, W. F., Wood, P., Fekete, B. M., Levy, M. A., & Watson, J. E. M. (2016). Sixteen years of change in the global terrestrial human footprint and implications for biodiversity conservation. *Nature Communications*, 7(1), 12558. <https://doi.org/10.1038/ncomms12558>

Venter, O., Sanderson, E. W., Magrach, A., Allan, J. R., Beher, J., Jones, K. R., Possingham, H. P., Laurance, W. F., Wood, P., Fekete, B. M., Levy, M. A., & Watson, J. E. M. (2017). *Data from: Sixteen years of change in the global terrestrial human footprint and implications for biodiversity conservation*. <https://doi.org/10.1038/ncomms12558>

- Global terrestrial Human Footprint maps for 1993 and 2009* (Version 2, p. 1720945220 bytes) [dataset]. [object Object]. <https://doi.org/10.5061/DRYAD.052Q5>
- Vickery, J. A., Ewing, S. R., Smith, K. W., Pain, D. J., Bairlein, F., Škorpilová, J., & Gregory, R. D. (2014). The decline of Afro-Palaearctic migrants and an assessment of potential causes. *Ibis*, 156(1), 1–22. <https://doi.org/10.1111/ibi.12118>
- Visser, M. E., Perdeck, A. C., Van BALEN, J. H., & Both, C. (2009). Climate change leads to decreasing bird migration distances. *Global Change Biology*, 15(8), 1859–1865. <https://doi.org/10.1111/j.1365-2486.2009.01865.x>
- Wauchope, H. S., Jones, J. P. G., Geldmann, J., Simmons, B. I., Amano, T., Blanco, D. E., Fuller, R. A., Johnston, A., Langendoen, T., Mundkur, T., Nagy, S., & Sutherland, W. J. (2022). Protected areas have a mixed impact on waterbirds, but management helps. *Nature*, 605(7908), 103–107. <https://doi.org/10.1038/s41586-022-04617-0>
- Wilcove, D. S., & Wikelski, M. (2008). Going, Going, Gone: Is Animal Migration Disappearing. *PLOS Biology*, 6(7), e188. <https://doi.org/10.1371/journal.pbio.0060188>
- Williams, B. A., Venter, O., Allan, J. R., Atkinson, S. C., Rehbein, J. A., Ward, M., Di Marco, M., Grantham, H. S., Ervin, J., Goetz, S. J., Hansen, A. J., Jantz, P., Pillay, R., Rodríguez-Buriticá, S., Supples, C., Virnig, A. L. S., & Watson, J. E. M. (2020). Change in Terrestrial Human Footprint Drives Continued Loss of Intact Ecosystems. *One Earth*, 3(3), 371–382. <https://doi.org/10.1016/j.oneear.2020.08.009>
- Wilson, S., Saracco, J. F., Krikun, R., Flockhart, D. T. T., Godwin, C. M., & Foster, K. R. (2018). Drivers of demographic decline across the annual cycle of a threatened migratory bird. *Scientific Reports*, 8(1), 7316. <https://doi.org/10.1038/s41598-018-25633-z>
- Zhang, W., Wei, J., & Xu, Y. (2023). Prioritizing global conservation of migratory birds over their migration network. *One Earth*, 6(10), 1340–1349. <https://doi.org/10.1016/j.oneear.2023.08.017>

## Discussion

The main objective of my thesis was to better understand principles governing the biogeography of food webs and the seasonal rearrangement of species and traits. To do so, I mainly focused on species' functional traits and used a macroecological approach, analyzing data on hundreds to thousands of species across continental scale. This large scale approach revealed novel generalities across communities, space, and time. In this discussion, I will present the key findings, their implications, limitations, and the future directions in three major themes of the thesis: trait-based predictions of trophic interactions, biogeography of food webs, and seasonal bird biogeography.

*Table 5: Summary of each chapter. The table presents the main question, summarizes methods, main findings and implications of each chapter.*

Chapter	Question	Methodology	Main Findings	Implications
1	How well we can predict interactions across large spatial scales and food webs ?	We used the food web of European tetrapods and species traits to train a series of Bayesian hierarchical linear models.	- Models calibrated with very few known interactions estimated the entire food web reasonably well. - Predictability of interactions was homogeneous in space. - Interactions of some species (e.g., generalist species) are not as well predicted.	- Trait relationships between interacting species are general within the European food web. - Trait-based models can make preliminary predictions to fill interaction data shortfalls.
2	How well trait-based interaction models can extrapolate to predict interactions	We trained models of trophic interactions on four vertebrate food webs (Canadian	- Trait-based models predict most interactions correctly in other food webs.	- Trait-based models can make first approximation of food webs in area with little to no

	and network properties in new ecological communities?	tundra, Serengeti, Pyrenees, and Europe) and evaluate how well these models predict each food web.	- Transferability decreased with increasing phylogenetic and environmental dissimilarity. - Network properties were less well predicted than interactions.	data. - Food web variation is mainly driven by variation in traits rather than by variation in trait relationships between interacting species. - Constraint beyond trait relationships are shaping food webs.
3	How bird traits explain birds degree to endure winter versus migrating to be exposed to similar climatic conditions across seasons?	We modelled variations in North American bird distributions and climate they experience across seasons as a function of their traits.	- Species traits explain migratory strategies with body mass better explaining movement in climatic space and the hand-wing index indicating movement in geographic space. - Species that tend to track climate conditions do not follow the classic latitudinal richness gradient in summer or the body mass-temperature relationship in winter, explaining some of the discrepancies in the literature	- The seasonal redistribution of bird and trait diversity across seasons will influence the functioning of communities. - Given changing seasonal dynamics, we need a better understanding of how birds interact with climate across their annual cycle to better anticipate the consequences of climate change.
4	How do seasonal level of human impact, protection, and bird traits explain the trends among North American bird species?	We combined seasonal distribution of North American birds to quantify seasonal level of stress and protection and relate them to	- Most bird communities declined between 19% and 39% since 1970 throughout North America. - Level of stress and protection during all seasons,	- Birds that are more tolerant during breeding and nonbreeding to human impacts decline less. - It is important to integrate spatiotemporal

		<p>estimated species trends between 1970 and 2017.</p>	<p>especially during migration, seems to be important drivers of bird trends.</p> <ul style="list-style-type: none"> <li>- Smaller birds, and birds that travel longer distances tend to decline more than average</li> </ul>	<p>variations in stressors, protection measures, and species traits into conservation efforts.</p>
--	--	--	---	--

## Trait-based predictions of trophic interactions

Trophic interactions, where a predator species feeds on a prey species, are constrained by neutral and niche processes (Morales-Castilla *et al.* 2015). First, a species can only feed on another if they co-occur (neutral constraint) which results from environmental filters, dispersal limitations, and species interactions. Second, a predator will feed on prey only if the prey's traits fits within the predator's trophic niche (niche constraint). This 'match' between the trophic niche of the predator and the traits of the prey is captured by trait-based trophic interaction models. The accuracy of these models depends on the generalities in the trait-interaction relationships of predators and prey across taxa and ecosystems.

### *Findings and implications*

The Eltonian Shortfall – “the lack of knowledge on species' interactions and these interactions' effects on individual survival and fitness” – is one of the main data gap in ecology (Hortal *et al.* 2015). Indeed, we lack interaction data for most taxa and ecosystems. Our findings of generalities within food web (Chapter 1) and across food webs (Chapter 2) indicate that trait-based predictions can provide first-order approximations of food webs. We showed that even with limited data, these models can fill gaps in incomplete food web or predict interactions in ecosystems with little to no data. However, trait-based

models tend to systematically overestimate the number of interactions, suggesting that predicted interactions are candidate interactions. These predictions can then be refined through geographic subsets, expert opinion, or additional data.

Protecting species interactions should enable a more functional approach to conservation for several reasons: (1) interactions provide ecosystems functions (Dehling & Stouffer 2018), (2) from interactions, we can infer the roles of species within ecosystems (Cirtwill *et al.* 2018), and (3) we can anticipate indirect consequences of managing species on other species in the ecosystems (McDonald-Madden *et al.* 2016). So far, this opportunity has still been underused, in part due to the lack of interaction data. The ability to predict interactions even in places with minimal data, as demonstrated in Chapters 1 and 2, should help overcome some of the barriers to integrating interaction data into conservation planning.

#### *Limitations and future directions*

1. Interaction strength, unlike the qualitative presence or absence of an interaction, is a continuous measurement quantifying the effect of one species on another (Berlow *et al.* 2004; Wootton & Emmerson 2005). The distribution of interaction strengths within a food web influence its stability (Ruiter *et al.* 1995) and functioning (Duffy 2002). Quantifying energy flows between species can also help anticipate cascading effects of perturbations (Bascompte *et al.* 2005). Yet, most trophic interactions model, including those used in Chapters 1 and 2, predict only the probability that a species could feed on another species. Future models that sought to quantify the effects of one species on another would be more useful.

2. Since we focused on trophic interactions between terrestrial vertebrates, our food webs do not include primary producers or invertebrates. These are, of course, important to understand how energy flows in ecosystems. Other models and traits could address this gap (Laigle *et al.* 2018; Li *et al.* 2023), but future research is needed to determine whether trait-interaction relationships are also generalizable for herbivory or invertebrate interactions.
3. In both Chapters 1 and 2, we used species-level traits, ignoring the potential role of intraspecific variation within and across regions on species interactions. Different environmental conditions and co-evolutionary dynamics can lead to shifts in trait-interaction relationships if the shifts in the traits of prey and predators are asymmetric (Gomulkiewicz *et al.* 2000). While it is reasonable to believe that intraspecific variation is less significant in regional and potential food webs compared to local and realized food webs, future studies should evaluate the extent and scale to which co-evolutionary dynamics influence trait-based predictions of trophic interactions and food webs.
4. When the primary goal is to fill knowledge gaps rather than ecological inferences, the future of trophic interaction predictions may lie on less ‘traditional’ methods. The rise of computing power and machine learning methods, like generative artificial intelligence, offers new opportunity to gather and infer interaction data from images, bioacoustics, texts, environmental DNA, or remote-sensing (Pichler & Hartig 2023). While machine learning algorithms hold great potential, they must be used with care as they can sometimes overfit, leading to low transferability, and may amplify existing biases in the data. However, using bias mitigation and regularization techniques can address these issues. These models have demonstrated the ability to transfer well to new conditions and perform effectively with limited and biased data (Pichler & Hartig 2023).

## Food web biogeography

Variation of food web structure along environmental gradients is driven by three processes: (1) changes in species (and trait) pool, (2) constraints acting at the food web level, and (3) changes in trait-interaction relationships (Poisot *et al.* 2015). The first process is the focus of filter, coexistence, and co-occurrence theories (Cazelles *et al.* 2016; Chesson 2000; Keddy 1992). This process results from the individual responses of species to the abiotic and biotic environment, and from dispersal limitations. The second process, food web-level constraints, include environmental, dynamic, and spatiotemporal factors. For example, primary productivity or ecosystem size might increase the feasible number of trophic level (Takimoto & Post 2013), while seasonality can create modules in food webs (McMeans *et al.* 2015), and stability might limit the number of possible interactions (Grilli *et al.* 2017). Finally, the third process, variation in trait-interaction relationships are the changes in the traits explaining why a predator might feed on a prey. For example, the body mass ratio between the consumer and resources might vary with temperature due to metabolic constraints (Brose *et al.* 2006). In Chapter 1 and 2, I investigated this third process among terrestrial vertebrates.

### *Findings and implications*

In Chapter 1, I showed that we were able to predict the food web of European terrestrial vertebrates relatively well using a fraction of all possible interactions. In Chapter 2, I found that a model calibrated in one ecosystem could predict most trophic interactions in other ecosystems given enough environmental and phylogenetic similarities. These results suggest that variation in trait-interaction

relationships is not the primary driver of food web variation. Instead, it seems like changes in species and trait distributions, as well as constraints on food web structure are more important.

These results are consistent with previous research finding generalities made on predator–prey body-size ratios within habitat, predator and prey types (Brose *et al.* 2006), and the trait-interaction relationships in soil invertebrates across three forest areas in Germany (Laigle *et al.* 2018).

Additionally, the generalities we found in trait-matching relationships within major taxonomic groups (order and class) in Chapter 1 and the decrease in model transferability with increasing phylogenetic similarity in Chapter 2, add to previous research indicating evolutionary conservation of species interactions (Gómez *et al.* 2010) and roles (Stouffer *et al.* 2012).

We also highlighted a major limitation of trait-based interaction models: they are unable to predict food web structures. This suggests that additional constraints are not captured beyond trait-interaction relationships. Factors such as area, spatial and temporal structures of the environment, for example, have been shown to shape ecological networks by creating modules and influencing specialization distributions (Galiana *et al.* 2018). Stability criteria also constrains the distribution of motifs, the number of interactions, and strength of these interactions (Borrelli 2015; Grilli *et al.* 2017). Finally, other environmental factors may influence the structure, such as primary productivity which constrains the number of trophic level and the diversity at each trophic levels.

#### *Limitations and future directions*

1. In Chapter 2, the scarcity of food web data limited us to only four food webs to work with. This small sample size resulted in large uncertainty for some of our model transferability results. This could be addressed when more fully resolved food web will become available. Having more ecological network data will also enable to test the generality of our transferability results for other ecosystems, taxonomic groups, and interaction types.
2. In Chapter 2, we found that trait-based interactions did not predict food web properties. This challenge mirrors issued that have been addressed in the field of species distribution modelling, where individual models predict species well, but not community level properties, such as species richness (Zurell *et al.* 2020). Methods have been developed to harness biases in higher level properties to correct predictions. For example, Leung *et al.* (2019) combined predicted species distributions and expected species richness to quantify species and spatial biases. They then correlated these biases with species traits and environmental covariates to improve predicted species distributions. Similar approaches integrating predictions of interactions and networks have the potential to provide better food web predictions (Isaac *et al.* 2020). By leveraging insights from other fields and adapting them to the context of food web predictions, we may develop more robust models, and better test the relative role of changes in species composition, food-web level-constraints, and trait-interaction relationships in food web variations.

## Seasonal bird biogeography

Every year, almost 20% of the world's bird species migrate, sometimes over thousands of kilometres, in response to seasonal variations in the environment (Kirby *et al.* 2008). This proportion of migratory

species raise to much higher numbers highly seasonal environments. This extraordinary adaptation to seasonality presumably evolved to maintain fidelity to breeding grounds, while escaping harsh conditions (Winger *et al.* 2019). Compared to other strategy, migration has the advantage to allow species to maintain relatively similar climate conditions throughout the annual cycle, but comes with the energetic cost of moving. In Chapter 3, I explore how traits explain how species balance the cost of sustaining different climate conditions versus the cost of moving in space.

Migration also means that species depend on many habitat, and their ability to move between these habitats. Migratory species are thus exposed to varying levels of stress and protection throughout their annual cycle (Runge *et al.* 2015; Rushing *et al.* 2017). This could explain why migratory birds experiencing declines at much higher rates than resident species (Wilcove & Wikelski 2008). In Chapter 4, I ask whether the level of human impact and protection at different seasons explain variations in trends among North American bird species. I also explored the traits of species that are especially vulnerable to decline.

### *Findings and implications*

Our results support the hypothesis the energetic trade-off between movement within their climatic niche and geographic space that migratory birds need to balance, and that their traits influence this. However, the causality of the relationship between traits and migratory strategy remains unclear. It is uncertain whether traits evolved to favor a particular migratory strategy, or if migratory strategies evolved in response to certain bird traits. With climate change altering seasonal dynamics and habitat

availability, traits that favor movement within climatic niche space over movement across geographic space may become more critical.

The bird communities showed a large variation in structure across the continent and seasons, both in terms of taxonomic diversity and also traits. It is therefore essential to understand the seasonal dynamics of bird communities to better protect, and anticipate changes in functioning. Some of the ecogeographical rules only stand for some species and seasons, giving insights into how those spatial structures might arise. For example, we found contrasting effects of seasons on the temperature-body size relationships among species that track climate conditions versus the ones that do not. This might indicate different constraints (heat dissipation versus heat conservation) driving body mass distributions in these two groups (Henry *et al.* 2023).

In Chapter 4, we highlighted the importance of the migration season in protecting species. Spatial planning research could enhance conservation strategies by comparing the prioritization of landscapes during different seasons, potentially identifying areas that provide greater protection for the full annual cycle of species (Runge *et al.* 2015). By considering the specific needs and vulnerabilities of migratory birds during their migration periods, conservation planners can identify and prioritize key habitats and stopover sites critical for their survival. This seasonal approach to spatial planning offers a comprehensive strategy for conserving migratory species across their entire annual cycle (Guo *et al.* 2024).

The findings from Chapters 3 and 4 highlight the necessity of studying the entire annual cycle of bird species for effective conservation. Conservation efforts must encompass the full migratory network to protect these populations, especially as migratory birds face global declines and are exposed to rapidly changing seasonal dynamics (Zhang *et al.* 2023). By addressing critical periods like staging and stopover sites, often overlooked in traditional conservation strategies, we can better protect migrating birds.

### *Limitations and future directions*

1. In Chapter 3, our findings suggest that bird traits play a significant role in shaping movement patterns in both geographic and climatic spaces, although the relationships between traits, migration distance, and climate niche space can be complex. Our results support the hypothesis that long-distance migrants seek more resource accessibility rather than climate conditions (Somveille *et al.* 2019). It also highlights the need to consider another type of ‘tracking’: resource-tracking. Future research could explore the relationship between traits and bird movement in geographic space, as well as within different dimensions of their niches, including Eltonian and Grinnellian niches.
2. Chapter 3 also highlighted a broader limitation within the concept of the ecological niche. Traditionally, the ecological niche is defined as the combination of biotic and abiotic factors that allow a species to maintain a positive growth rate (Chase & Leibold 2003; Holt 2009). However, in a seasonal environment, these factors vary over time in a predictable way. Additionally, the responses and impacts of species can also vary to these biotic and abiotic factors as the species enters different stages of their annual cycle. By better integrating

seasonality into niche concepts and models, we can more accurately anticipate species' responses to changing climate conditions and seasonal dynamics.

3. In Chapter 4, we found contrasting effects of human footprint during different seasons on species trends. Higher human footprint was associated with declines during the migration season, but with increasing species trends during the breeding and wintering seasons.

Understanding how these contrasting effects arises to understand the drivers of bird decline.

One possible explanation is that the human footprint, as a composite of various human stressors, obscures more intricate interactions between individual stressors (such as light pollution and human density) at different seasons and species trends. Birds have distinct requirements during each season, and their vulnerability to individual stressors likely varies accordingly (Rushing *et al.* 2017). Additionally, Serratosa *et al* (2024) found that mortality did not exhibit a linear relationship with human impact, underscoring the potential for more complex relationships between human activities and bird populations.

## Litterature cited

- Bascompte, J., Melián, C.J. & Sala, E. (2005). Interaction strength combinations and the overfishing of a marine food web. *Proceedings of the National Academy of Sciences*, 102, 5443–5447.
- Berlow, E.L., Neutel, A.-M., Cohen, J.E., Ruiter, P.C.D., Ebenman, B., Emmerson, M., et al. (2004). Interaction strengths in food webs: issues and opportunities. *Journal of Animal Ecology*, 73, 585–598.
- Borrelli, J.J. (2015). Selection against instability: stable subgraphs are most frequent in empirical food webs. *Oikos*, 124, 1583–1588.
- Brose, U., Jonsson, T., Berlow, E.L., Warren, P., Banasek-Richter, C., Bersier, L.-F., et al. (2006). Consumer–Resource Body-Size Relationships in Natural Food Webs. *Ecology*, 87, 2411–2417.
- Cazelles, K., Araújo, M.B., Mouquet, N. & Gravel, D. (2016). A theory for species co-occurrence in interaction networks. *Theor Ecol*, 9, 39–48.
- Chase, J.M. & Leibold, M.A. (2003). *Ecological Niches: Linking Classical and Contemporary Approaches*. University of Chicago Press.
- Chesson, P. (2000). Mechanisms of Maintenance of Species Diversity. *Annual Review of Ecology and Systematics*, 31, 343–366.
- Cirtwill, A.R., Dalla Riva, G.V., Gaiarsa, M.P., Bimler, M.D., Cagua, E.F., Coux, C., et al. (2018). A review of species role concepts in food webs. *Food Webs*, 16, e00093.
- Dehling, D.M. & Stouffer, D.B. (2018). Bringing the Eltonian niche into functional diversity. *Oikos*, 127, 1711–1723.
- Duffy, J.E. (2002). Biodiversity and ecosystem function: the consumer connection. *Oikos*, 99, 201–219.
- Galiana, N., Lurgi, M., Claramunt-López, B., Fortin, M.-J., Leroux, S., Cazelles, K., et al. (2018). The spatial scaling of species interaction networks. *Nat Ecol Evol*, 2, 782–790.

- Gómez, J.M., Verdú, M. & Perfectti, F. (2010). Ecological interactions are evolutionarily conserved across the entire tree of life. *Nature*, 465, 918–921.
- Gomulkiewicz, R., Thompson, J.N., Holt, R.D., Nuismer, S.L. & Hochberg, M.E. (2000). Hot Spots, Cold Spots, and the Geographic Mosaic Theory of Coevolution. *The American Naturalist*, 156, 156–174.
- Grilli, J., Adorisio, M., Suweis, S., Barabás, G., Banavar, J.R., Allesina, S., *et al.* (2017). Feasibility and coexistence of large ecological communities. *Nat Commun*, 8, 14389.
- Guo, F., Buler, J.J., Smolinsky, J.A. & Wilcove, D.S. (2024). Seasonal patterns and protection status of stopover hotspots for migratory landbirds in the eastern United States. *Current Biology*, 34, 235-244.e3.
- Henry, E., Santini, L., Huijbregts, M.A.J. & Benítez-López, A. (2023). Unveiling the environmental drivers of intraspecific body size variation in terrestrial vertebrates. *Global Ecology and Biogeography*, 32, 267–280.
- Holt, R.D. (2009). Bringing the Hutchinsonian niche into the 21st century: Ecological and evolutionary perspectives. *PNAS*, 106, 19659–19665.
- Hortal, J., de Bello, F., Diniz-Filho, J.A.F., Lewinsohn, T.M., Lobo, J.M. & Ladle, R.J. (2015). Seven Shortfalls that Beset Large-Scale Knowledge of Biodiversity. *Annu. Rev. Ecol. Evol. Syst.*, 46, 523–549.
- Isaac, N.J.B., Jarzyna, M.A., Keil, P., Damblay, L.I., Boersch-Supan, P.H., Browning, E., *et al.* (2020). Data Integration for Large-Scale Models of Species Distributions. *Trends in Ecology & Evolution*, 35, 56–67.
- Keddy, P.A. (1992). Assembly and response rules: two goals for predictive community ecology. *Journal of Vegetation Science*, 3, 157–164.

- Kirby, J.S., Stattersfield, A.J., Butchart, S.H.M., Evans, M.I., Grimmett, R.F.A., Jones, V.R., *et al.* (2008). Key conservation issues for migratory land- and waterbird species on the world's major flyways. *Bird Conservation International*, 18, S49–S73.
- Laigle, I., Aubin, I., Digel, C., Brose, U., Boulangeat, I. & Gravel, D. (2018). Species traits as drivers of food web structure. *Oikos*, 127, 316–326.
- Leung, B., Hudgins, E.J., Potapova, A. & Ruiz-Jaen, M.C. (2019). A new baseline for countrywide  $\alpha$ -diversity and species distributions: illustration using >6,000 plant species in Panama. *Ecological Applications*, 29, e01866.
- Li, J., Luo, M., Wang, S., Gauzens, B., Hirt, M.R., Rosenbaum, B., *et al.* (2023). A size-constrained feeding-niche model distinguishes predation patterns between aquatic and terrestrial food webs. *Ecology Letters*, 26, 76–86.
- McDonald-Madden, E., Sabbadin, R., Game, E.T., Baxter, P.W.J., Chadès, I. & Possingham, H.P. (2016). Using food-web theory to conserve ecosystems. *Nat Commun*, 7, 1–8.
- McMeans, B.C., McCann, K.S., Humphries, M., Rooney, N. & Fisk, A.T. (2015). Food Web Structure in Temporally-Forced Ecosystems. *Trends in Ecology & Evolution*, 30, 662–672.
- Morales-Castilla, I., Matias, M.G., Gravel, D. & Araújo, M.B. (2015). Inferring biotic interactions from proxies. *Trends in Ecology & Evolution*, 30, 347–356.
- Pichler, M. & Hartig, F. (2023). Machine learning and deep learning—A review for ecologists. *Methods in Ecology and Evolution*, 14, 994–1016.
- Poisot, T., Stouffer, D.B. & Gravel, D. (2015). Beyond species: why ecological interaction networks vary through space and time. *Oikos*, 124, 243–251.
- Ruiter, P.C. de, Neutel, A.-M. & Moore, J.C. (1995). Energetics, Patterns of Interaction Strengths, and Stability in Real Ecosystems. *Science*, 269, 1257–1260.

Runge, C.A., Watson, J.E.M., Butchart, S.H.M., Hanson, J.O., Possingham, H.P. & Fuller, R.A. (2015).

Protected areas and global conservation of migratory birds. *Science*, 350, 1255–1258.

Rushing, C.S., Hostetler, J.A., Sillett, T.S., Marra, P.P., Rotenberg, J.A. & Ryder, T.B. (2017). Spatial and temporal drivers of avian population dynamics across the annual cycle. *Ecology*, 98, 2837–2850.

Serratosa, J., Oppel, S., Rotics, S., Santangeli, A., Butchart, S.H.M., Cano-Alonso, L.S., *et al.* (2024).

Tracking data highlight the importance of human-induced mortality for large migratory birds at a flyway scale. *Biological Conservation*, 293, 110525.

Somveille, M., Manica, A. & Rodrigues, A.S.L. (2019). Where the wild birds go: explaining the differences in migratory destinations across terrestrial bird species. *Ecography*, 42, 225–236.

Stouffer, D.B., Sales-Pardo, M., Sirer, M.I. & Bascompte, J. (2012). Evolutionary Conservation of Species' Roles in Food Webs. *Science*, 335, 1489–1492.

Takimoto, G. & Post, D.M. (2013). Environmental determinants of food-chain length: a meta-analysis. *Ecol Res*, 28, 675–681.

Wilcove, D.S. & Wikelski, M. (2008). Going, Going, Gone: Is Animal Migration Disappearing. *PLOS Biology*, 6, e188.

Winger, B.M., Auteri, G.G., Pegan, T.M. & Weeks, B.C. (2019). A long winter for the Red Queen: rethinking the evolution of seasonal migration. *Biological Reviews*, 94, 737–752.

Wootton, J.T. & Emmerson, M. (2005). Measurement of Interaction Strength in Nature. *Annual Review of Ecology, Evolution, and Systematics*, 36, 419–444.

Zhang, W., Wei, J. & Xu, Y. (2023). Prioritizing global conservation of migratory birds over their migration network. *One Earth*, 6, 1340–1349.

Zurell, D., Zimmermann, N.E., Gross, H., Baltensweiler, A., Sattler, T. & Wüest, R.O. (2020). Testing species assemblage predictions from stacked and joint species distribution models. *Journal of Biogeography*, 47, 101–113.

# Appendices

## A. Supplementary material for chapter 1

Addressing the Eltonian shortfall with trait-based interaction models

*Dominique Caron, Luigi Maiorano, Wilfried Thuiller, Laura Pollock*

### **Table of Content:**

**Appendix S1 – Amphibian body mass imputation:** Relationship between amphibians body length and body mass

**Appendix S2 - Model runs and potential scale reduction factor:** List of all models with the scale reduction factor for each parameter to check for convergence

**Appendix S3 - Predictive model sensitivity analyses:** Sensitivity analyses testing the influence of the number of absences used for calibration, the procedure for sampling non-interactions, the size of the validation dataset and the random sampling for validation dataset.

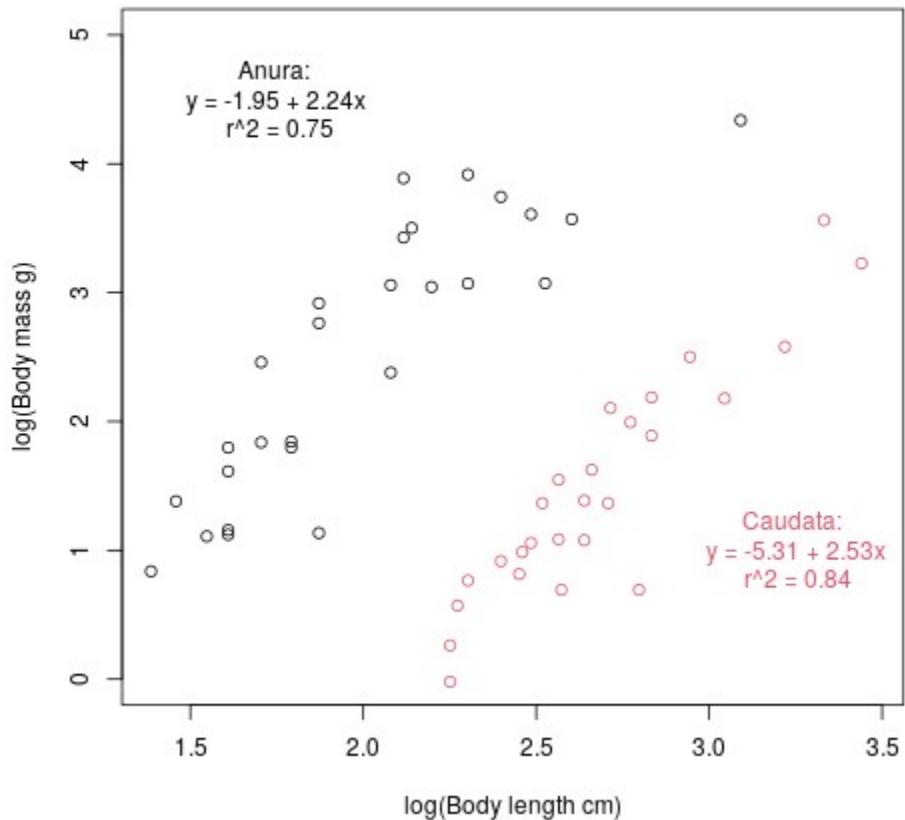
**Appendix S4 - Predictor Importance:** Parameter estimates and quantiles used to measured the importance of each predictor.

**Appendix S5 - Miscalibration results:** Results of the miscalibration analysis

**Appendix S6 - Bioregional models and food webs:** Comparison of the parameters inferred for the models calibrated in each bioregion, and the predictive performance of each of theses model to predict the food web of every bioregion.

**Appendix S7 - : Additional performance metrics:** Figure 1 and 2 of the paper showed with additional performance metrics.

## Appendix S1: Amphibian body mass imputation



*Figure S1: Relationship between body length and body mass for amphibian species. We used the equations shown in this figure to impute body mass from body length for each order (Anura and Caudata) separately. The vertical axis is the log of body mass in g and horizontal axis is the log of body length in cm.*

## Appendix S2: Model runs and potential scale reduction factor

*Table S2: List of all models with the potential scale reduction factor. In order, the columns are: a unique Id for the model, the origin database of interactions used for calibration, the number*

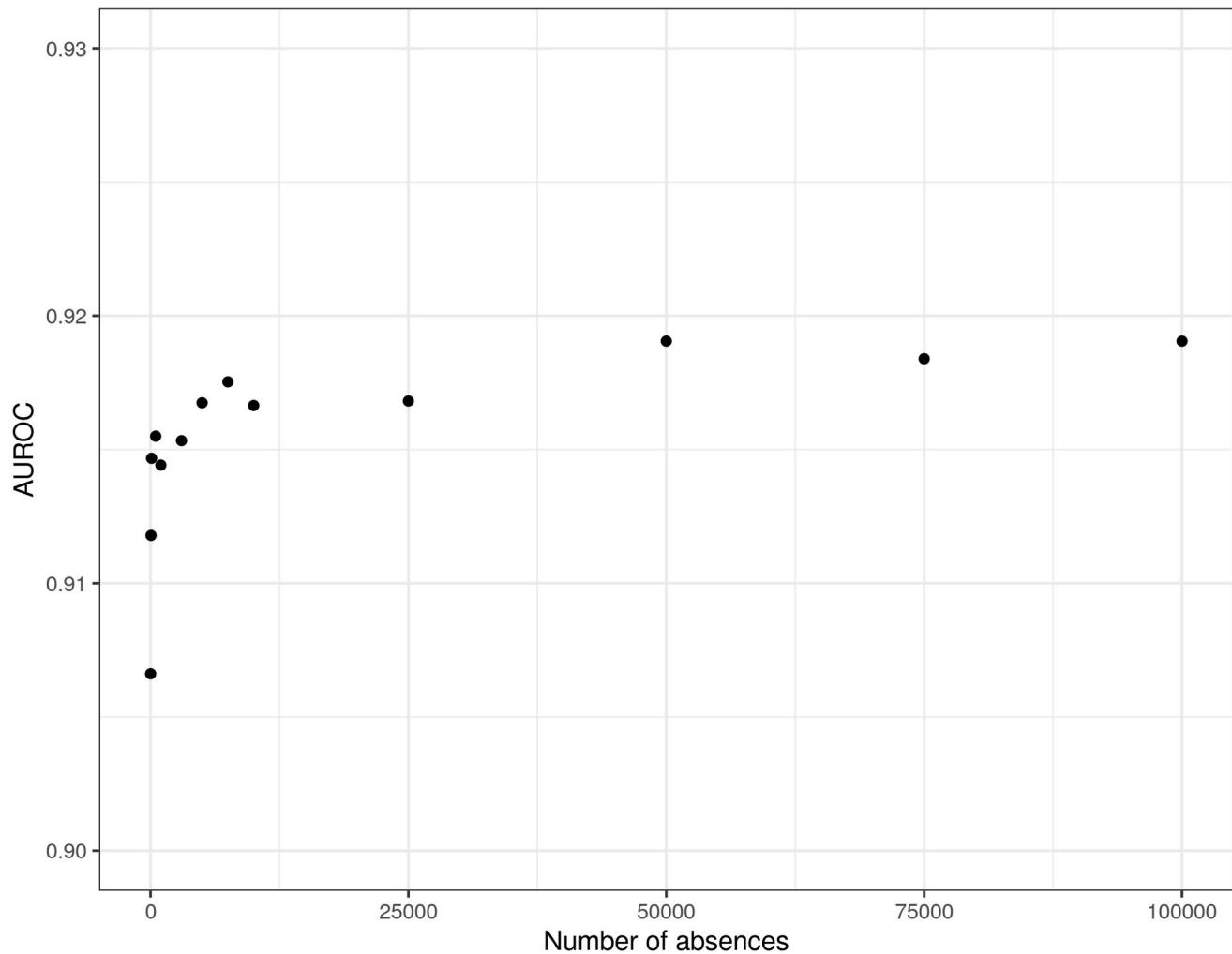
RunID	TrainingDB	Interactions	Intercept	BodyMassPred	OpportunisticPred	HuntingPred	BrowserPred	GrazerPred	BodyMassPrey	DietMatch	ActivityTimeMatch	HabitatMatch	BodyMassMatch
1	globi	2	1.001	1.003	1.001	1.000	1.001	1.000	1.000	1.001	1.001	1.000	1.001
2	globi	2	1.002	1.001	1.002	1.003	1.000	1.000	1.000	1.001	1.000	1.002	1.000
3	globi	2	1.000	1.001	1.001	1.001	1.001	1.001	1.001	1.002	1.000	1.001	1.000
4	globi	2	1.000	1.002	1.000	1.001	1.000	1.001	1.000	1.001	1.001	1.001	1.001
5	globi	2	1.001	1.001	1.000	1.000	1.000	1.001	1.001	1.000	1.001	1.004	1.000
6	globi	5	1.001	1.002	1.001	1.001	1.000	1.001	1.000	1.001	1.000	1.001	1.001
7	globi	5	1.000	1.000	1.000	1.000	1.000	1.002	1.000	1.000	1.002	1.002	1.001
8	globi	5	1.000	1.000	1.000	1.001	1.001	1.001	1.000	1.001	1.002	1.001	1.000
9	globi	5	1.001	1.001	1.000	1.001	1.003	1.001	1.001	1.001	1.000	1.001	1.000
10	globi	5	1.000	1.000	1.001	1.000	1.001	1.002	1.003	1.002	1.002	1.000	1.001
11	globi	10	1.000	1.000	1.001	1.000	1.000	1.000	1.000	1.000	1.001	1.001	1.000
12	globi	10	1.000	1.000	1.002	1.000	1.000	1.000	1.000	1.000	1.000	1.001	1.000
13	globi	10	1.002	1.000	1.001	1.001	1.000	1.000	1.000	1.000	1.001	1.000	1.000
14	globi	10	1.000	1.001	1.000	1.001	1.000	1.001	1.000	1.001	1.001	1.000	1.000
15	globi	10	1.001	1.001	1.000	1.000	1.000	1.002	1.000	1.000	1.000	1.001	1.000
16	globi	25	1.001	1.000	1.001	1.000	1.001	1.000	1.000	1.001	1.001	1.008	1.001
17	globi	25	1.000	1.000	1.000	1.001	1.001	1.001	1.000	1.001	1.000	1.000	1.000
18	globi	25	1.001	1.002	1.001	1.001	1.002	1.001	1.002	1.000	1.000	1.001	1.000
19	globi	25	1.001	1.000	1.001	1.001	1.001	1.002	1.003	1.004	1.008	1.003	1.001
20	globi	25	1.002	1.004	1.000	1.001	1.001	1.000	1.000	1.002	1.004	1.001	1.001
21	globi	50	1.000	1.000	1.001	1.000	1.001	1.001	1.000	1.002	1.000	1.003	1.000
22	globi	50	1.000	1.001	1.000	1.000	1.000	1.002	1.001	1.001	1.001	1.001	1.001
23	globi	50	1.005	1.005	1.001	1.004	1.010	1.002	1.003	1.001	1.000	1.017	1.002
24	globi	50	1.002	1.003	1.001	1.001	1.004	1.001	1.005	1.001	1.000	1.002	1.005
25	globi	50	1.001	1.007	1.002	1.003	1.001	1.001	1.007	1.002	1.006	1.000	1.004
26	globi	100	1.001	1.003	1.001	1.001	1.002	1.011	1.002	1.000	1.001	1.000	1.001
27	globi	100	1.003	1.000	1.002	1.000	1.007	1.001	1.000	1.001	1.001	1.003	1.001
28	globi	100	1.001	1.005	1.001	1.002	1.001	1.003	1.003	1.000	1.004	1.002	1.005
29	globi	100	1.001	1.006	1.001	1.002	1.000	1.001	1.005	1.001	1.002	1.001	1.005
30	globi	100	1.000	1.000	1.001	1.000	1.001	1.000	1.001	1.000	1.002	1.001	1.000
31	globi	250	1.003	1.000	1.001	1.002	1.001	1.002	1.000	1.000	1.003	1.003	1.000
32	globi	250	1.002	1.003	1.001	1.002	1.000	1.003	1.002	1.002	1.003	1.002	1.002
33	globi	250	1.001	1.001	1.000	1.002	1.001	1.000	1.000	1.000	1.002	1.000	1.001
34	globi	250	1.000	1.004	1.001	1.000	1.001	1.002	1.002	1.001	1.000	1.001	1.002
35	globi	250	1.000	1.000	1.000	1.000	1.002	1.001	1.000	1.000	1.008	1.001	1.000
36	globi	291	1.004	1.001	1.000	1.004	1.004	1.010	1.002	1.000	1.001	1.005	1.000
37	globi	291	1.003	1.001	1.000	1.002	1.000	1.000	1.001	1.001	1.000	1.002	1.001
38	globi	291	1.001	1.001	1.002	1.000	1.001	1.000	1.001	1.001	1.001	1.002	1.000
39	globi	291	1.002	1.007	1.000	1.002	1.000	1.006	1.005	1.002	1.007	1.005	1.007
40	globi	291	1.002	1.000	1.001	1.002	1.001	1.001	1.001	1.001	1.009	1.001	1.001
41	metaweb	2	1.000	1.000	1.001	1.001	1.004	1.000	1.000	1.000	1.001	1.002	1.000
42	metaweb	2	1.000	1.000	1.002	1.002	1.000	1.000	1.001	1.000	1.006	1.000	1.001
43	metaweb	2	1.003	1.001	1.000	1.002	1.003	1.001	1.002	1.001	1.002	1.003	1.001
44	metaweb	2	1.001	1.001	1.001	1.001	1.000	1.003	1.002	1.001	1.000	1.002	1.000
45	metaweb	2	1.001	1.001	1.002	1.000	1.000	1.000	1.003	1.001	1.000	1.005	1.001
46	metaweb	5	1.000	1.001	1.003	1.000	1.001	1.001	1.001	1.000	1.000	1.001	1.000
47	metaweb	5	1.002	1.001	1.001	1.000	1.000	1.002	1.000	1.000	1.001	1.003	1.000
48	metaweb	5	1.000	1.001	1.002	1.000	1.000	1.002	1.000	1.000	1.002	1.001	1.001
49	metaweb	5	1.000	1.002	1.001	1.000	1.000	1.003	1.002	1.002	1.001	1.003	1.001
50	metaweb	5	1.001	1.001	1.001	1.001	1.000	1.000	1.000	1.001	1.001	1.003	1.001
51	metaweb	10	1.000	1.000	1.001	1.000	1.002	1.001	1.001	1.001	1.001	1.001	1.000
52	metaweb	10	1.000	1.001	1.000	1.000	1.000	1.000	1.000	1.000	1.001	1.000	1.000
53	metaweb	10	1.000	1.000	1.001	1.000	1.001	1.001	1.000	1.001	1.000	1.002	1.000
54	metaweb	10	1.000	1.001	1.000	1.001	1.000	1.001	1.000	1.000	1.002	1.001	1.000
55	metaweb	10	1.001	1.000	1.000	1.000	1.000	1.001	1.000	1.001	1.000	1.000	1.000
56	metaweb	25	1.001	1.000	1.000	1.001	1.002	1.000	1.001	1.000	1.001	1.001	1.000
57	metaweb	25	1.001	1.001	1.001	1.000	1.002	1.000	1.000	1.000	1.000	1.001	1.000
58	metaweb	25	1.001	1.001	1.000	1.001	1.000	1.000	1.000	1.001	1.000	1.000	1.001
59	metaweb	25	1.000	1.001	1.000	1.001	1.001	1.000	1.000	1.000	1.000	1.000	1.001
60	metaweb	25	1.000	1.000	1.000	1.001	1.000	1.000	1.000	1.001	1.000	1.000	1.001
61	metaweb	50	1.001	1.001	1.001	1.000	1.002	1.001	1.000	1.001	1.000	1.000	1.002
62	metaweb	50	1.001	1.002	1.003	1.001	1.000	1.001	1.002	1.000	1.000	1.000	1.001
63	metaweb	50	1.000	1.002	1.000	1.000	1.000	1.000	1.000	1.001	1.001	1.000	1.000
64	metaweb	50	1.001	1.002	1.001	1.001	1.000	1.002	1.001	1.000	1.001	1.000	1.002
65	metaweb	50	1.000	1.003	1.001	1.000	1.000	1.002	1.002	1.001	1.000	1.001	1.002
66	metaweb	100	1.000	1.001	1.000	1.000	1.000	1.001	1.000	1.000	1.002	1.000	1.000
67	metaweb	100	1.003	1.009	1.001	1.001	1.002	1.001	1.004	1.001	1.002	1.000	1.007
68	metaweb	100	1.000	1.001	1.000	1.000	1.001	1.000	1.000	1.000	1.001	1.001	1.001
69	metaweb	100	1.001	1.001	1.001	1.000	1.000	1.000	1.000	1.001	1.001	1.000	1.002
70	metaweb	100	1.000	1.003	1.001	1.001	1.001	1.000	1.001	1.001	1.001	1.020	1.001

**Table S2 (continued): List of all models with the potential scale reduction factor. In order, the columns are: a unique Id for the model, the origin database of interactions used for calibration, the number**

RunID	TrainingDB	Interactions	Intercept	BodyMassPred	OpportunisticPred	HuntingPred	BrowserPred	GrazerPred	BodyMassPrey	DietMatch	ActivityTimeMatch	HabitatMatch	BodyMassMatch
71	metaweb	250	1.000	1.000	1.000	1.000	1.001	1.000	1.000	1.000	1.000	1.000	1.000
72	metaweb	250	1.001	1.000	1.002	1.001	1.000	1.002	1.002	1.001	1.002	1.002	1.001
73	metaweb	250	1.001	1.001	1.001	1.001	1.001	1.001	1.000	1.000	1.000	1.001	1.001
74	metaweb	250	1.001	1.009	1.001	1.001	1.001	1.001	1.005	1.002	1.003	1.001	1.009
75	metaweb	250	1.000	1.001	1.001	1.000	1.002	1.000	1.001	1.000	1.002	1.000	1.003
76	metaweb	500	1.002	1.002	1.000	1.001	1.002	1.002	1.001	1.000	1.000	1.001	1.001
77	metaweb	500	1.001	1.003	1.001	1.000	1.002	1.008	1.005	1.000	1.001	1.002	1.005
78	metaweb	500	1.000	1.002	1.002	1.000	1.001	1.003	1.002	1.001	1.001	1.000	1.003
79	metaweb	500	1.002	1.004	1.001	1.002	1.001	1.006	1.003	1.001	1.003	1.004	1.003
80	metaweb	500	1.001	1.002	1.000	1.000	1.000	1.002	1.003	1.000	1.001	1.002	1.001
81	metaweb	750	1.000	1.002	1.000	1.001	1.002	1.000	1.001	1.001	1.003	1.001	1.002
82	metaweb	750	1.001	1.003	1.001	1.002	1.001	1.003	1.001	1.001	1.000	1.002	1.001
83	metaweb	750	1.000	1.001	1.000	1.000	1.001	1.003	1.003	1.002	1.001	1.002	1.001
84	metaweb	750	1.002	1.004	1.001	1.001	1.000	1.004	1.002	1.000	1.005	1.001	1.003
85	metaweb	750	1.000	1.002	1.000	1.000	1.001	1.001	1.000	1.000	1.000	1.000	1.000
86	metaweb	1000	1.000	1.002	1.004	1.001	1.002	1.004	1.002	1.002	1.005	1.002	1.003
87	metaweb	1000	1.001	1.002	1.001	1.001	1.001	1.000	1.001	1.000	1.000	1.001	1.001
88	metaweb	1000	1.000	1.003	1.000	1.000	1.002	1.004	1.004	1.000	1.002	1.002	1.002
89	metaweb	1000	1.002	1.002	1.001	1.001	1.001	1.001	1.001	1.001	1.000	1.002	1.001
90	metaweb	1000	1.000	1.002	1.000	1.000	1.002	1.001	1.004	1.001	1.017	1.002	1.003
91	metaweb	1500	1.004	1.002	1.001	1.002	1.002	1.001	1.002	1.000	1.001	1.002	1.002
92	metaweb	1500	1.002	1.002	1.000	1.001	1.001	1.001	1.001	1.001	1.002	1.003	1.001
93	metaweb	1500	1.001	1.000	1.000	1.000	1.002	1.000	1.000	1.000	1.001	1.003	1.000
94	metaweb	1500	1.003	1.000	1.001	1.001	1.001	1.004	1.002	1.000	1.000	1.007	1.000
95	metaweb	1500	1.000	1.004	1.001	1.000	1.003	1.008	1.004	1.001	1.009	1.001	1.004
96	metaweb	2000	1.004	1.000	1.002	1.001	1.002	1.002	1.005	1.008	1.006	1.001	1.001
97	metaweb	2000	1.000	1.001	1.002	1.000	1.000	1.001	1.001	1.001	1.000	1.001	1.001
98	metaweb	2000	1.002	1.002	1.004	1.001	1.002	1.010	1.002	1.002	1.000	1.002	1.001
99	metaweb	2000	1.001	1.001	1.000	1.001	1.001	1.006	1.000	1.001	1.000	1.000	1.000
100	metaweb	2000	1.002	1.001	1.002	1.002	1.001	1.000	1.000	1.000	1.003	1.001	1.001
101	metaweb	3000	1.002	1.001	1.001	1.002	1.000	1.001	1.000	1.001	1.007	1.001	1.000
102	metaweb	3000	1.004	1.001	1.002	1.002	1.002	1.001	1.001	1.001	1.002	1.007	1.001
103	metaweb	3000	1.000	1.008	1.002	1.000	1.000	1.000	1.003	1.026	1.003	1.001	1.003
104	metaweb	3000	1.002	1.007	1.001	1.002	1.004	1.000	1.006	1.001	1.011	1.003	1.006
105	metaweb	3000	1.000	1.006	1.005	1.000	1.001	1.001	1.002	1.001	1.002	1.001	1.008
106	metaweb	4000	1.000	1.001	1.001	1.001	1.000	1.000	1.000	1.001	1.000	1.001	1.001
107	metaweb	4000	1.000	1.000	1.004	1.002	1.002	1.002	1.001	1.008	1.007	1.007	1.001
108	metaweb	4000	1.000	1.001	1.011	1.001	1.006	1.016	1.002	1.006	1.007	1.002	1.004
109	metaweb	4000	1.009	1.005	1.033	1.005	1.003	1.017	1.010	1.003	1.041	1.001	1.013
110	metaweb	4000	1.004	1.001	1.005	1.002	1.002	1.004	1.001	1.001	1.000	1.000	1.001
111	metaweb	5000	1.003	1.005	1.006	1.003	1.002	1.002	1.002	1.000	1.007	1.005	1.002
112	metaweb	5000	1.003	1.004	1.003	1.001	1.001	1.000	1.001	1.001	1.000	1.003	1.003
113	metaweb	5000	1.003	1.003	1.015	1.001	1.004	1.025	1.008	1.005	1.017	1.046	1.010
114	metaweb	5000	1.000	1.005	1.003	1.002	1.000	1.001	1.005	1.001	1.001	1.015	1.004
115	metaweb	5000	1.000	1.005	1.027	1.000	1.004	1.007	1.003	1.026	1.040	1.049	1.003
Master	metaweb	7157	1.002	1.002	1.007	1.001	1.007	1.018	1.005	1.001	1.012	1.021	1.007
General	metaweb	71417	1.001	1.000	NA	NA	NA	NA	1.001	1.001	1.012	1.011	1.000
Anura	metaweb	916	1.002	1.001	NA	NA	NA	NA	1.000	1.001	1.004	1.002	
Caudata	metaweb	811	1.002	1.001	NA	NA	NA	NA	1.003	1.000	1.001	1.006	
Gaviiformes	metaweb	44	1.004	1.002	NA	NA	NA	NA	1.001	1.002	1.000	1.002	
Podicipediformes	metaweb	109	1.001	1.000	NA	NA	NA	NA	1.001	1.000	1.001	1.001	
Procellariiformes	metaweb	3	1.004	1.006	NA	NA	NA	NA	1.001	1.003	1.005	1.000	
Pelecaniformes	metaweb	10	1.001	1.001	NA	NA	NA	NA	1.000	1.000	1.000	1.002	
Ciconiiformes	metaweb	2599	1.001	1.001	NA	NA	NA	NA	1.002	1.000	1.006	1.003	
Anseriformes	metaweb	1139	1.021	1.020	NA	NA	NA	NA	1.024	1.000	1.001	1.000	
Falconiformes	metaweb	8069	1.002	1.001	NA	NA	NA	NA	1.002	1.001	1.002	1.001	
Galliformes	metaweb	493	1.002	1.003	NA	NA	NA	NA	1.005	NA	1.000	1.003	
Gruiformes	metaweb	1203	1.003	1.003	NA	NA	NA	NA	1.003	1.000	1.001	1.002	
Charadriiformes	metaweb	4407	1.005	1.002	NA	NA	NA	NA	1.009	1.001	1.005	1.015	
Cuculiformes	metaweb	87	1.000	1.001	NA	NA	NA	NA	1.002	1.000	1.001	1.000	
Strigiformes	metaweb	3915	1.003	1.002	NA	NA	NA	NA	1.002	1.000	1.001	1.001	
Coraciiformes	metaweb	753	1.000	1.000	NA	NA	NA	NA	1.003	1.001	1.000	1.003	
Piciformes	metaweb	129	1.000	1.000	NA	NA	NA	NA	1.001	1.000	1.000	1.000	
Passeriformes	metaweb	8390	1.001	1.002	NA	NA	NA	NA	1.002	1.001	1.001	1.008	
Rodentia	metaweb	1644	1.001	1.001	NA	NA	NA	NA	1.001	1.001	1.001	1.004	
Cetartiodactyla	metaweb	491	1.011	1.009	NA	NA	NA	NA	1.003	1.002	1.015	1.020	
Carnivora	metaweb	8837	1.001	1.000	NA	NA	NA	NA	1.000	1.002	1.002	1.004	
Eulipotyphla	metaweb	1589	1.004	1.005	NA	NA	NA	NA	1.008	1.000	1.002	1.001	
Chiroptera	metaweb	20	1.000	1.001	NA	NA	NA	NA	1.001	1.001	1.000	1.001	
Squamata	metaweb	25759	1.002	1.001	NA	NA	NA	NA	1.001	1.001	1.001	1.001	

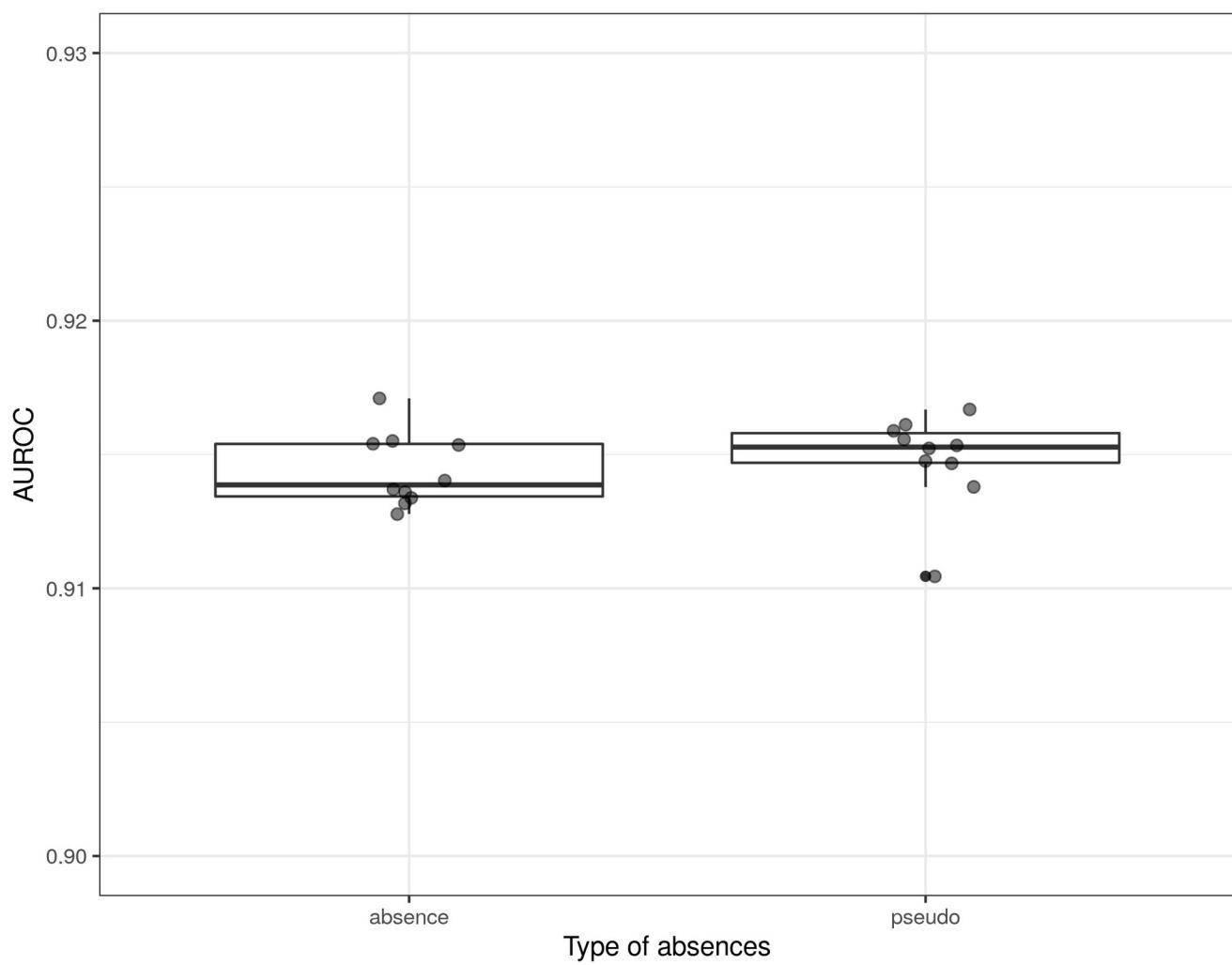
### Appendix S3: Predictive model sensitivity analyses

AUROC estimate vs Number of absences in training dataset



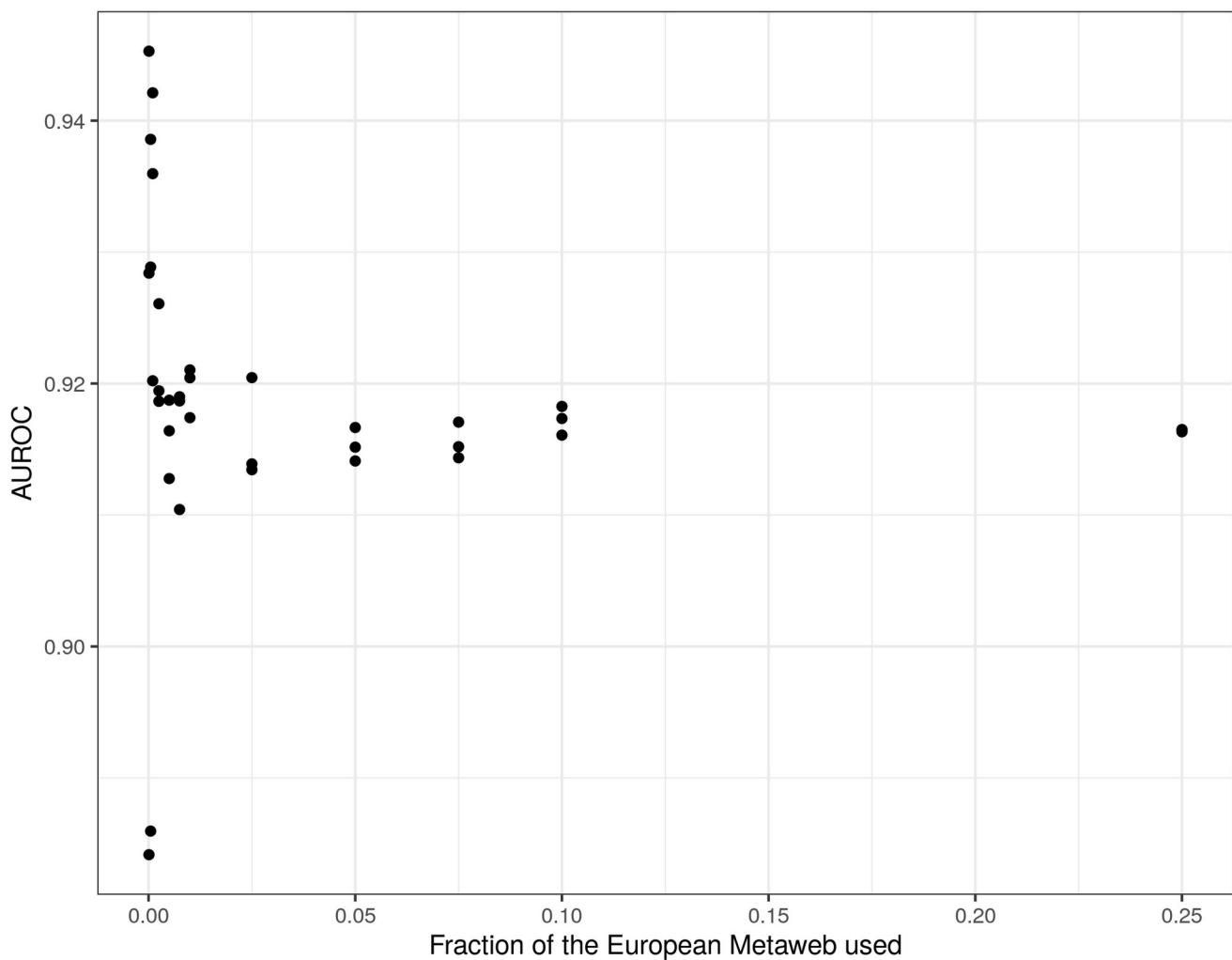
*Figure S3.1: Effect of the number of absences sampled for the training dataset on the predictive performance of the model. Predictive performance was measured on the same 1% of the entire European Metaweb as validation dataset. The number of absences varies from 200 to 100 000, and we kept the number of interactions in the training dataset constant at 300.*

### AUROC estimate vs Procedure or absence selection



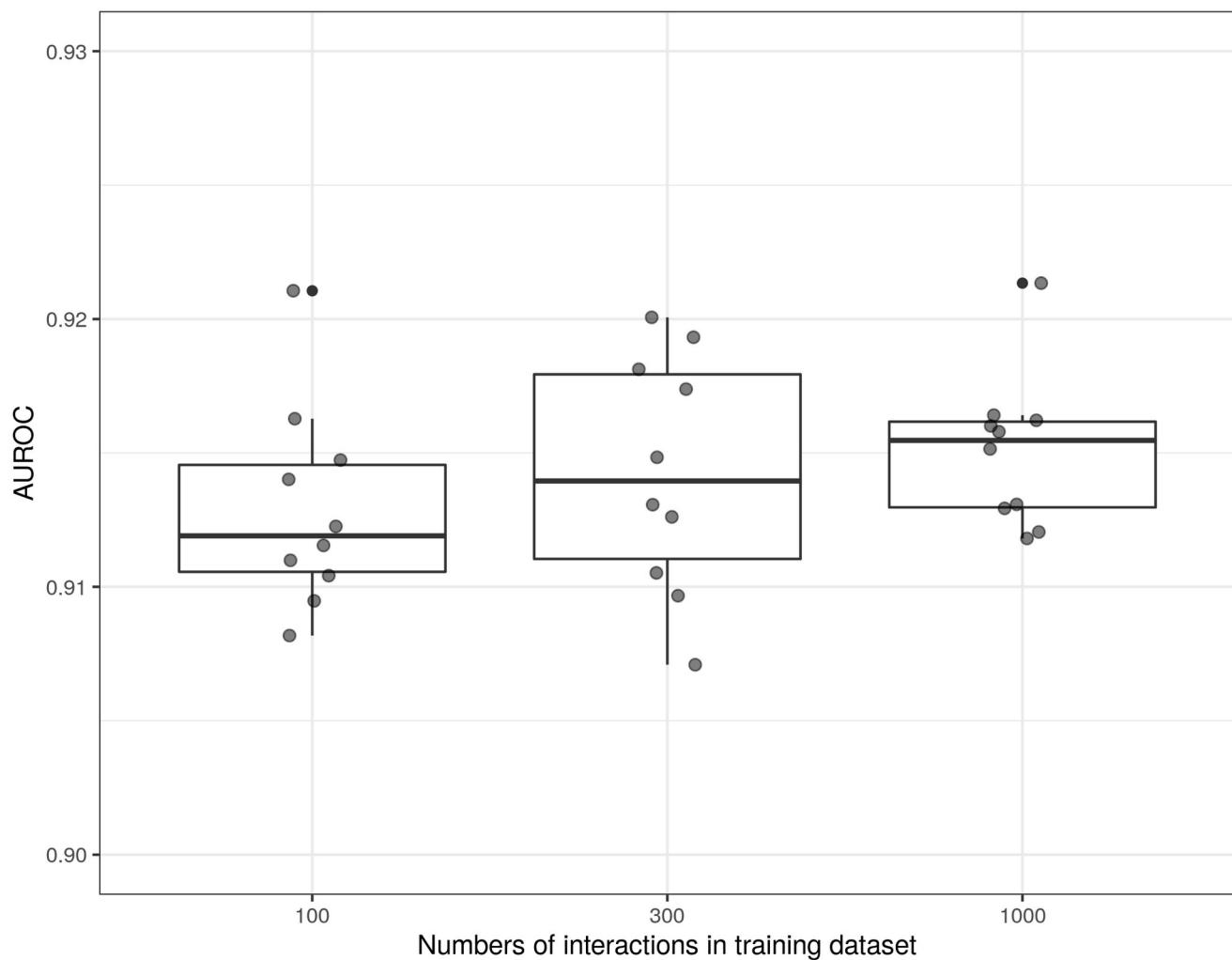
*Figure S3.2: Effect of the sampling procedures of absences of interaction on the predictive performance of the models. We measured performance using the same 1% of the European Metaweb as validation dataset. We sampled 300 interactions and 10 000 non-interactions from the European Metaweb to train the models. The non-interactions were either absences (0 in the dataset) or pseudo-absences (random pairs of species; i.e. background data).*

### AUROC estimate vs Size of the validation dataset



*Figure S3.3: Effect of the size of the validation dataset on the precision of the predictive performance estimate. We sampled 3 times different sizes of validation dataset (from 0.01% to 25% of the entire European Metaweb) to measure the predictive performance of a model trained from 300 interactions and 10 000 non-interactions sampled from the European Metaweb.*

### AUROC estimate vs Different validation datasets



*Figure S3.4: Effect of random sampling the validation dataset on the predictive performance estimate. We trained 3 models using 3 different training dataset sizes, and measured their performance on 10 randomly sampled validation datasets from the European Metaweb, all of which represent 1% of the entire European Metaweb.*

## Appendix S4: Predictor Importance

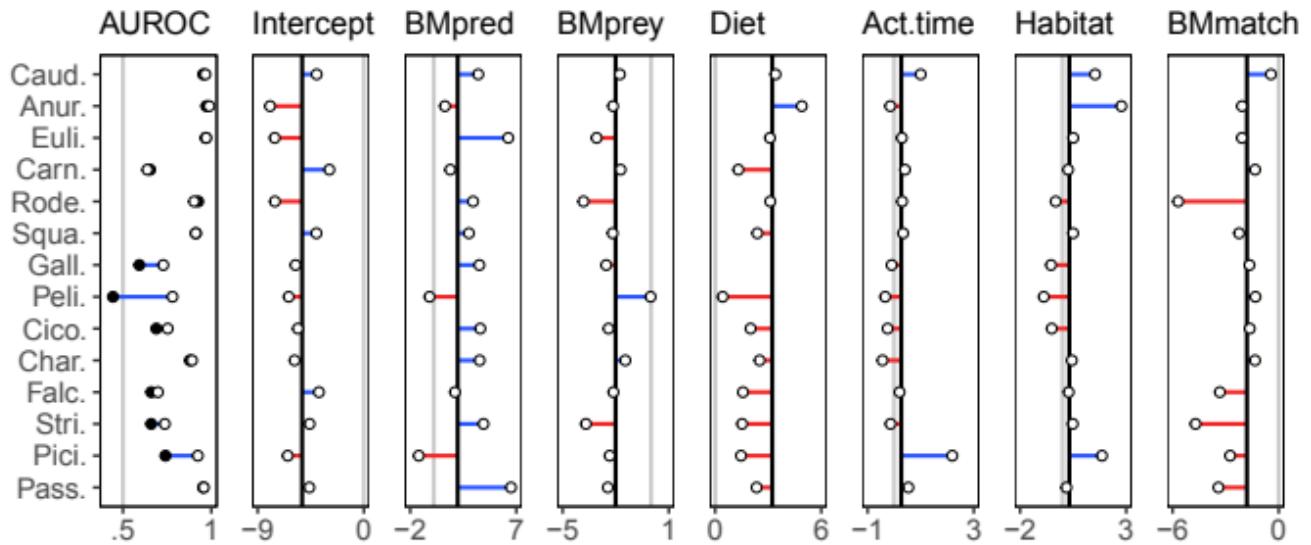
*Table S4.1: Summary statistics for parameters inferred for the predictive model trained on the entire European Metaweb. All non-binary predictors were scaled. The predictors are: the log of the body mass of the predator (continuous), the four categories of foraging behaviours (binary), the log of the body mass of the prey (continuous), the match between the diet of the predator and the category of prey (binary), the match between the activity time of the interacting species (continuous), the match between the nesting habitat of the interacting species (continuous), and the squared difference of the log of the body mass of the predator and the log of the body mass of the prey (continuous).*

	Mean	SD	Naive SE	Time-series SE
Intercept	-4.5660	0.0117	0.0002	0.0004
Predator body mass	2.0788	0.0199	0.0003	0.0008
ForBehav.Opportunistic	0.6073	0.0116	0.0002	0.0006
ForBehav.Hunting	0.9256	0.0189	0.0003	0.0011
ForBehav.Browser	-0.0827	0.0120	0.0002	0.0014
ForBehav.Grazer	-0.1856	0.0155	0.0002	0.0021
Prey body mass	-1.9946	0.0172	0.0003	0.0006
Diet match	2.2923	0.0114	0.0002	0.0003
Activity time match	0.2559	0.0092	0.0001	0.0009
Nesting habitat match	0.3141	0.0092	0.0001	0.0006
Body mass difference	-1.7453	0.0213	0.0003	0.0009

*Table S4.2: Quantiles of the parameters inferred for the predictive model trained on the entire European Metaweb. All non-binary predictors were scaled*

	2.5%	25%	50%	75%	97.5%
Intercept	-4.589	-4.574	-4.566	-4.558	-4.544
Predator body mass	2.041	2.065	2.079	2.092	2.118
ForBehav.Opportunistic	0.584	0.600	0.607	0.615	0.631
ForBehav.Hunting	0.890	0.913	0.925	0.938	0.963
ForBehav.Browser	-0.109	-0.090	-0.082	-0.075	-0.061
ForBehav.Grazer	-0.215	-0.197	-0.186	-0.176	-0.156
Prey body mass	-2.028	-2.006	-1.995	-1.983	-1.960
Diet match	2.270	2.285	2.292	2.300	2.315
Activity time match	0.239	0.249	0.256	0.262	0.274
Nesting habitat match	0.296	0.308	0.314	0.320	0.334
Body mass difference	-1.787	-1.759	-1.746	-1.730	-1.703

## Appendix S5: Miscalibration results



*Figure S5: Miscalibration of the general model for each group of predators. The groups are the same as in Figure 2. The left panel shows the differences between the predictive performance of the general model (black dots) and the group-specific models (white dots). The other panels show the differences between the parameters inferred (i.e. the miscalibration) for the general model (black lines) and each the group-specific model (white dots). From left to right, the parameters are: the intercept, the linear coefficients associated with the body mass of the predator, the body mass of the prey, the match between the diet of the predator and the type of prey, the similarity in activity time, the similarity in nesting habitat, and the difference between the body mass of the predator and the prey.*

## Appendix S6: Bioregional models and food webs

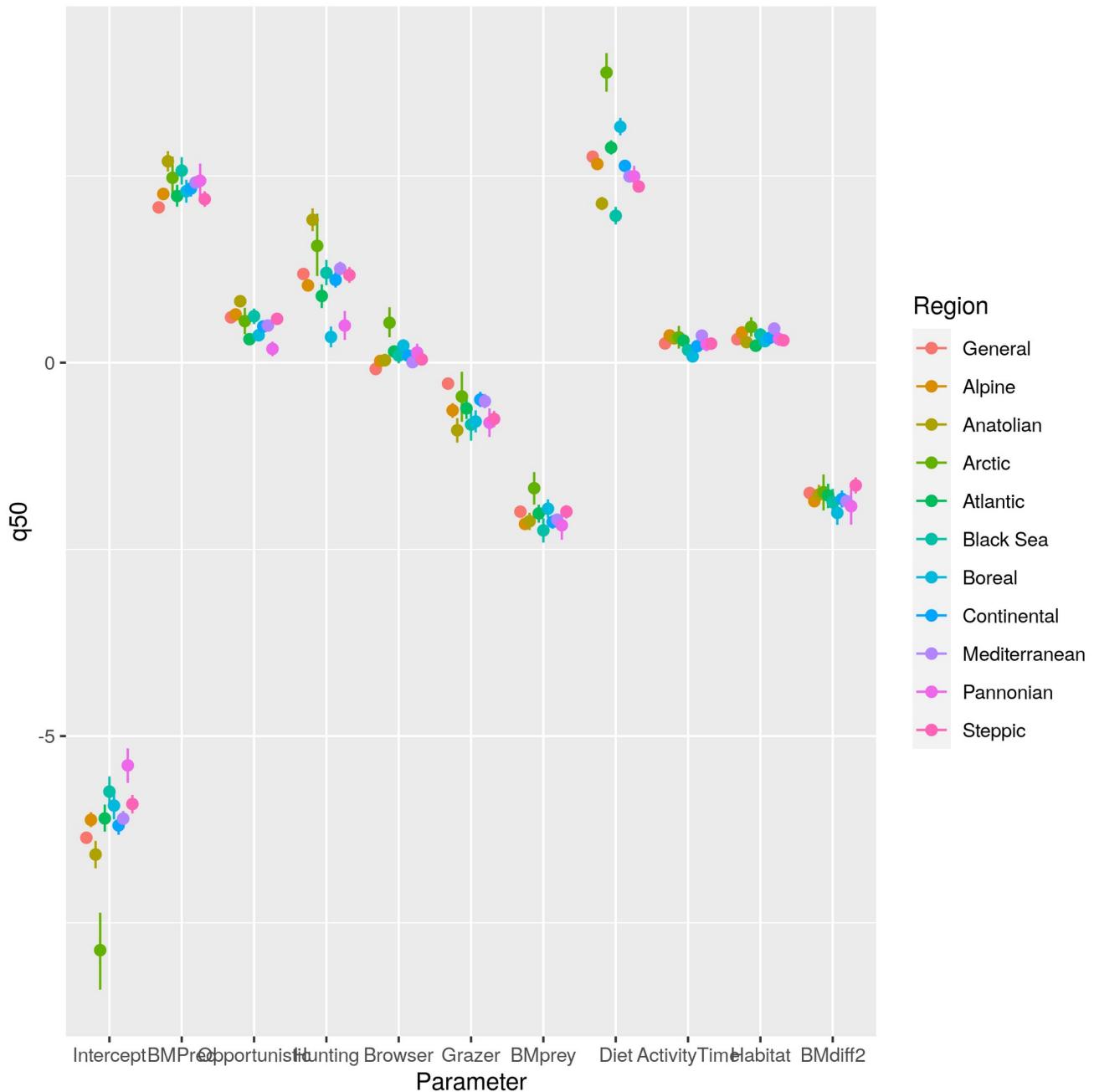


Figure S6: Comparison of the bioregional food webs models. The red points are the inferred parameter for the general model (calibrated on the entire food web), and the other points are the inferred parameters of the models calibrated on each bioregional food web.

*Table S6: Transferability of predictive models across bio-regions. Performance metrics of the models calibrated on each bio-regions (Training FW) for predicting the food webs of each bio-regions (Testing FW). Performanc metrics are: AUC (area under the receiver operating curve), TSS (true skill statistics), TPR (true positive rate), TNR (true negative rate), FPR (false positive rate), and FNR (false negative rate).*

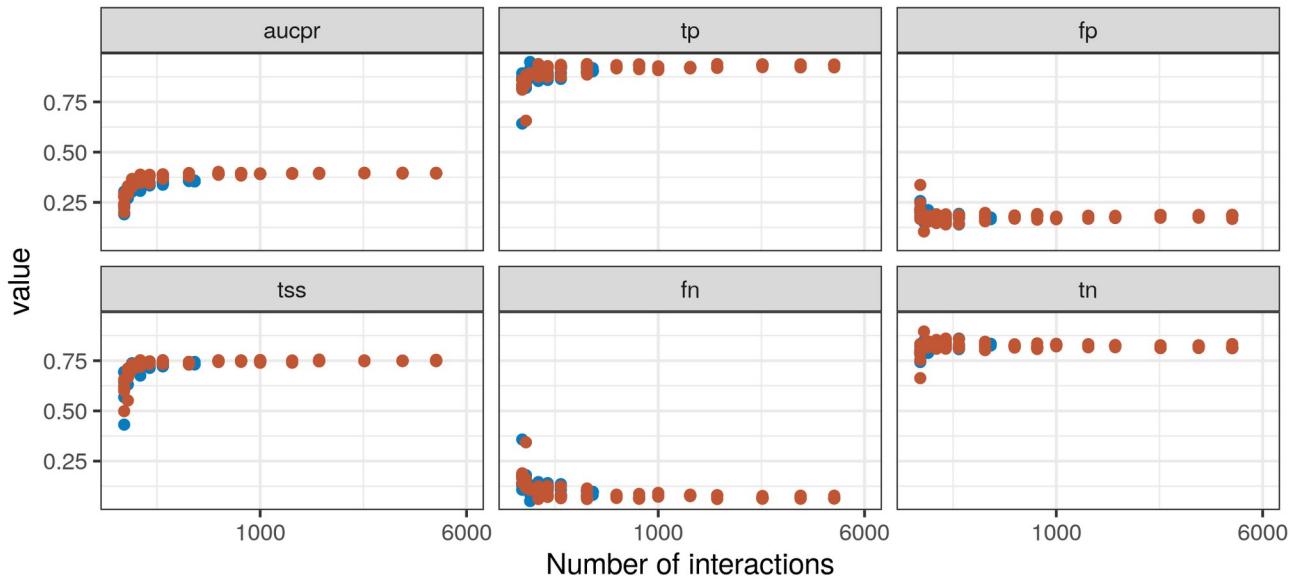
Training FW	Testing FW	AUC	TSS	TPR	TNR	FPR	FNR
general	general	0.92	0.73	0.91	0.82	0.18	0.09
general	alpine	0.92	0.73	0.92	0.81	0.19	0.08
general	anatolian	0.92	0.71	0.91	0.81	0.19	0.09
general	arctic	0.94	0.76	0.96	0.81	0.19	0.04
general	atlantic	0.91	0.70	0.91	0.79	0.21	0.09
general	blackSea	0.90	0.69	0.90	0.80	0.20	0.10
general	boreal	0.91	0.73	0.95	0.78	0.22	0.05
general	continental	0.91	0.71	0.91	0.80	0.20	0.09
general	mediterranean	0.91	0.71	0.91	0.80	0.20	0.09
general	pannonian	0.89	0.68	0.90	0.77	0.23	0.10
general	steppic	0.90	0.69	0.90	0.79	0.21	0.10
alpine	general	0.92	0.73	0.90	0.83	0.17	0.10
alpine	alpine	0.92	0.74	0.91	0.83	0.17	0.09
alpine	anatolian	0.92	0.72	0.90	0.82	0.18	0.10
alpine	arctic	0.94	0.79	0.97	0.83	0.17	0.03
alpine	atlantic	0.91	0.71	0.91	0.80	0.20	0.09
alpine	blackSea	0.90	0.70	0.89	0.81	0.19	0.11
alpine	boreal	0.90	0.73	0.93	0.80	0.20	0.07
alpine	continental	0.91	0.71	0.90	0.81	0.19	0.10
alpine	mediterranean	0.91	0.70	0.89	0.82	0.18	0.11
alpine	pannonian	0.89	0.69	0.91	0.78	0.22	0.09
alpine	steppic	0.90	0.70	0.89	0.81	0.19	0.11
anatolian	general	0.92	0.73	0.91	0.82	0.18	0.09
anatolian	alpine	0.92	0.73	0.92	0.81	0.19	0.08
anatolian	anatolian	0.92	0.74	0.93	0.81	0.19	0.07
anatolian	arctic	0.94	0.77	0.96	0.81	0.19	0.04
anatolian	atlantic	0.90	0.71	0.92	0.79	0.21	0.08
anatolian	blackSea	0.90	0.70	0.91	0.80	0.20	0.09
anatolian	boreal	0.90	0.71	0.93	0.79	0.21	0.07
anatolian	continental	0.91	0.72	0.92	0.80	0.20	0.08
anatolian	mediterranean	0.91	0.71	0.90	0.80	0.20	0.10
anatolian	pannonian	0.89	0.68	0.91	0.77	0.23	0.09
anatolian	steppic	0.91	0.72	0.92	0.80	0.20	0.08
arctic	general	0.92	0.73	0.88	0.84	0.16	0.12
arctic	alpine	0.91	0.72	0.88	0.83	0.17	0.12
arctic	anatolian	0.91	0.70	0.87	0.83	0.17	0.13
arctic	arctic	0.94	0.81	0.96	0.85	0.15	0.04
arctic	atlantic	0.90	0.69	0.89	0.80	0.20	0.11
arctic	blackSea	0.90	0.66	0.85	0.81	0.19	0.15
arctic	boreal	0.90	0.71	0.91	0.80	0.20	0.09
arctic	continental	0.90	0.69	0.88	0.81	0.19	0.12
arctic	mediterranean	0.91	0.69	0.87	0.82	0.18	0.13
arctic	pannonian	0.89	0.67	0.89	0.78	0.22	0.11
arctic	steppic	0.90	0.67	0.86	0.81	0.19	0.14
atlantic	general	0.92	0.72	0.91	0.81	0.19	0.09
atlantic	alpine	0.92	0.72	0.92	0.80	0.20	0.08
atlantic	anatolian	0.92	0.74	0.94	0.80	0.20	0.06
atlantic	arctic	0.94	0.74	0.97	0.77	0.23	0.03
atlantic	atlantic	0.91	0.69	0.93	0.76	0.24	0.07
atlantic	blackSea	0.90	0.68	0.90	0.79	0.21	0.10
atlantic	boreal	0.90	0.71	0.95	0.76	0.24	0.05
atlantic	continental	0.91	0.71	0.93	0.78	0.22	0.07
atlantic	mediterranean	0.91	0.70	0.91	0.79	0.21	0.09
atlantic	pannonian	0.89	0.64	0.89	0.75	0.25	0.11
atlantic	steppic	0.90	0.68	0.90	0.78	0.22	0.10

*Table S6(continued): Transferability of predictive models across bio-regions. Performance metrics of the models calibrated on each bio-regions for predicting the food webs of each bio-regions.*

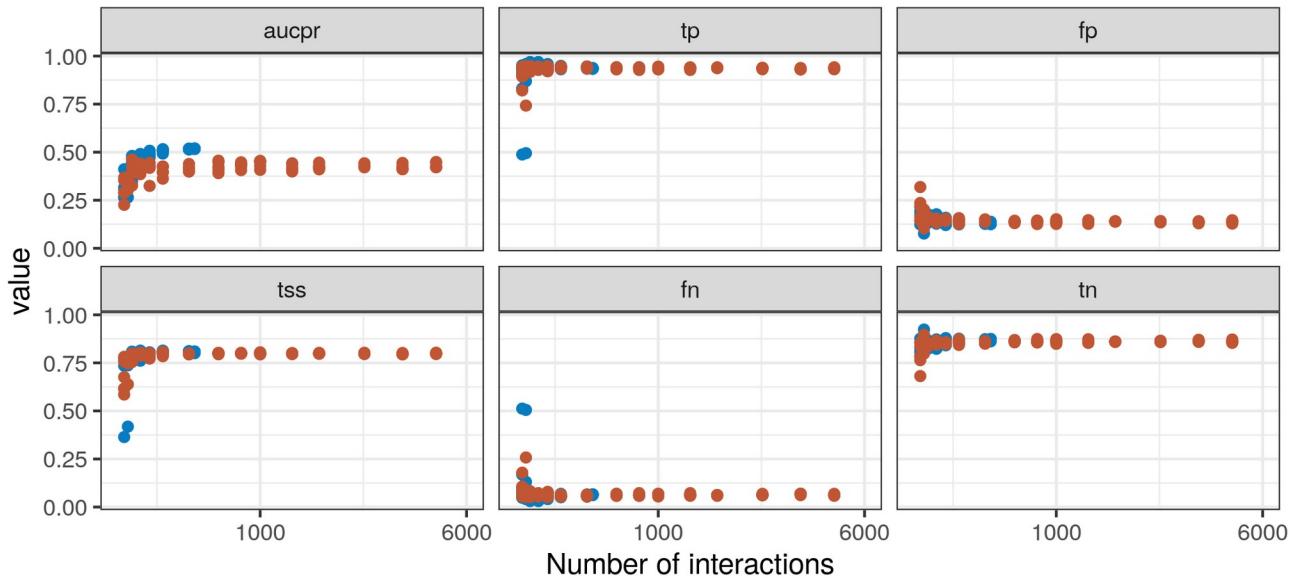
Training FW	Testing FW	AUC	TSS	TPR	TNR	FPR	FNR
blackSea	general	0.92	0.72	0.93	0.80	0.20	0.07
blackSea	alpine	0.91	0.73	0.94	0.80	0.20	0.06
blackSea	anatolian	0.92	0.71	0.92	0.79	0.21	0.08
blackSea	arctic	0.93	0.73	0.97	0.76	0.24	0.03
blackSea	atlantic	0.90	0.69	0.92	0.77	0.23	0.08
blackSea	blackSea	0.90	0.70	0.92	0.78	0.22	0.08
blackSea	boreal	0.91	0.72	0.95	0.77	0.23	0.05
blackSea	continental	0.91	0.71	0.93	0.78	0.22	0.07
blackSea	mediterranean	0.91	0.70	0.92	0.79	0.21	0.08
blackSea	pannonian	0.90	0.69	0.92	0.77	0.23	0.08
blackSea	steppic	0.90	0.70	0.93	0.78	0.22	0.07
boreal	general	0.91	0.72	0.87	0.85	0.15	0.13
boreal	alpine	0.92	0.72	0.88	0.84	0.16	0.12
boreal	anatolian	0.92	0.70	0.86	0.84	0.16	0.14
boreal	arctic	0.94	0.81	0.95	0.86	0.14	0.05
boreal	atlantic	0.90	0.71	0.89	0.82	0.18	0.11
boreal	blackSea	0.90	0.69	0.86	0.82	0.18	0.14
boreal	boreal	0.91	0.74	0.93	0.82	0.18	0.07
boreal	continental	0.91	0.72	0.90	0.82	0.18	0.10
boreal	mediterranean	0.91	0.70	0.87	0.83	0.17	0.13
boreal	pannonian	0.89	0.69	0.89	0.80	0.20	0.11
boreal	steppic	0.90	0.69	0.87	0.82	0.18	0.13
continental	general	0.92	0.72	0.91	0.81	0.19	0.09
continental	alpine	0.92	0.73	0.93	0.80	0.20	0.07
continental	anatolian	0.92	0.72	0.92	0.80	0.20	0.08
continental	arctic	0.93	0.76	0.98	0.78	0.22	0.02
continental	atlantic	0.91	0.71	0.93	0.77	0.23	0.07
continental	blackSea	0.90	0.70	0.92	0.79	0.21	0.08
continental	boreal	0.91	0.73	0.96	0.77	0.23	0.04
continental	continental	0.91	0.72	0.93	0.79	0.21	0.07
continental	mediterranean	0.91	0.71	0.92	0.80	0.20	0.08
continental	pannonian	0.89	0.67	0.91	0.76	0.24	0.09
continental	steppic	0.91	0.70	0.92	0.78	0.22	0.08
mediterranean	general	0.92	0.72	0.93	0.80	0.20	0.07
mediterranean	alpine	0.92	0.73	0.93	0.80	0.20	0.07
mediterranean	anatolian	0.92	0.72	0.93	0.79	0.21	0.07
mediterranean	arctic	0.94	0.73	0.99	0.74	0.26	0.01
mediterranean	atlantic	0.90	0.69	0.93	0.77	0.23	0.07
mediterranean	blackSea	0.90	0.70	0.91	0.78	0.22	0.09
mediterranean	boreal	0.91	0.72	0.96	0.76	0.24	0.04
mediterranean	continental	0.91	0.71	0.93	0.77	0.23	0.07
mediterranean	mediterranean	0.91	0.71	0.92	0.79	0.21	0.08
mediterranean	pannonian	0.89	0.66	0.91	0.75	0.25	0.09
mediterranean	steppic	0.90	0.69	0.93	0.77	0.23	0.07
pannonian	general	0.91	0.72	0.89	0.83	0.17	0.11
pannonian	alpine	0.91	0.72	0.90	0.82	0.18	0.10
pannonian	anatolian	0.91	0.72	0.89	0.82	0.18	0.11
pannonian	arctic	0.93	0.75	0.96	0.79	0.21	0.04
pannonian	atlantic	0.90	0.70	0.91	0.79	0.21	0.09
pannonian	blackSea	0.90	0.69	0.89	0.80	0.20	0.11
pannonian	boreal	0.91	0.72	0.93	0.79	0.21	0.07
pannonian	continental	0.91	0.72	0.92	0.80	0.20	0.08
pannonian	mediterranean	0.91	0.70	0.89	0.81	0.19	0.11
pannonian	pannonian	0.89	0.67	0.89	0.77	0.23	0.11
pannonian	steppic	0.90	0.69	0.88	0.80	0.20	0.12
steppic	general	0.92	0.73	0.93	0.80	0.20	0.07
steppic	alpine	0.91	0.73	0.94	0.80	0.20	0.06
steppic	anatolian	0.92	0.74	0.94	0.80	0.20	0.06
steppic	arctic	0.93	0.74	0.97	0.77	0.23	0.03
steppic	atlantic	0.90	0.69	0.93	0.76	0.24	0.07
steppic	blackSea	0.90	0.72	0.93	0.79	0.21	0.07
steppic	boreal	0.91	0.73	0.96	0.77	0.23	0.04
steppic	continental	0.91	0.72	0.94	0.78	0.22	0.06
steppic	mediterranean	0.91	0.70	0.92	0.78	0.22	0.08
steppic	pannonian	0.89	0.68	0.93	0.75	0.25	0.07
steppic	steppic	0.90	0.71	0.93	0.78	0.22	0.07

## Appendix S7: Additional performance metrics

### Predicting the Metaweb

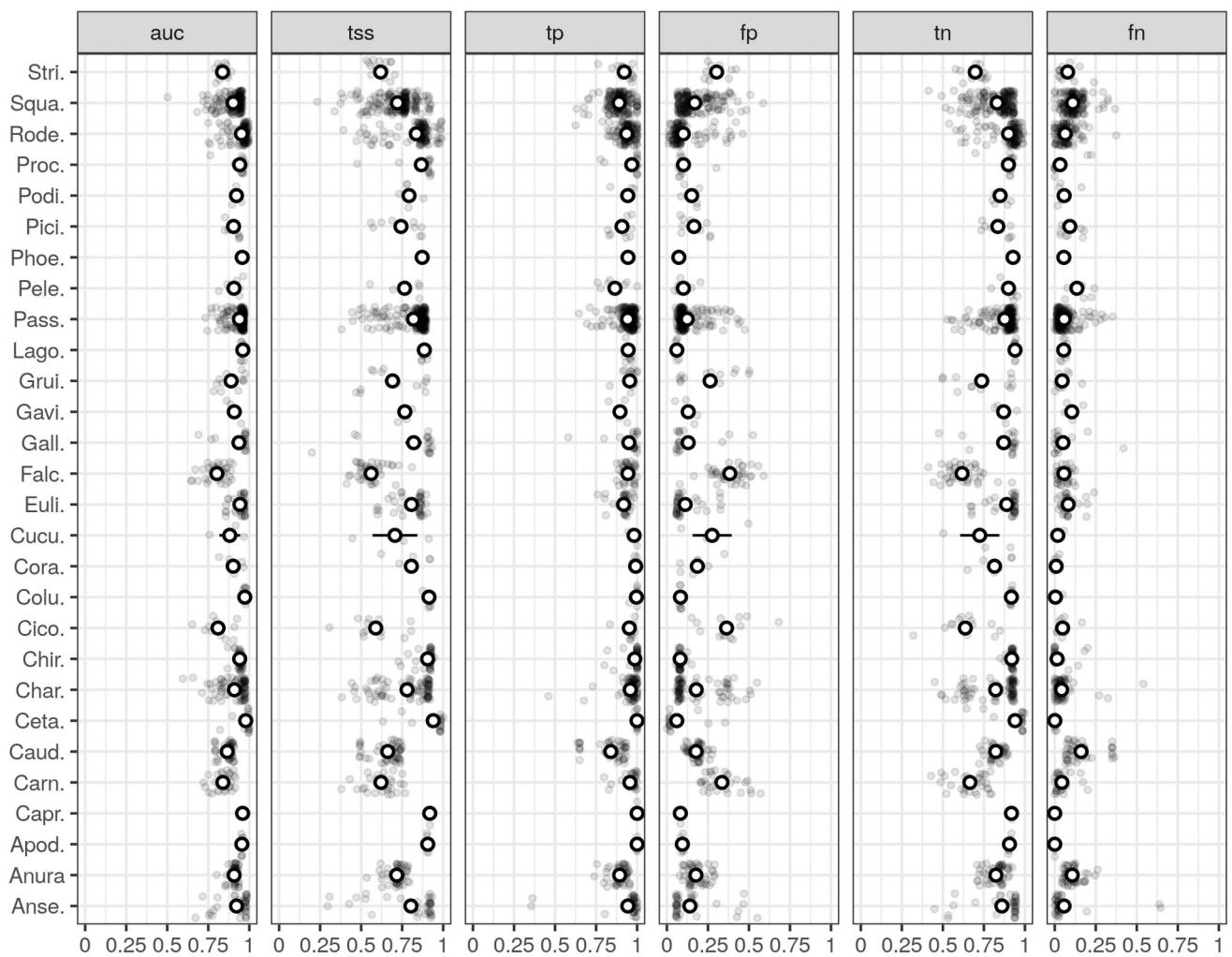


### Predicting Globi



*Figure S7.1: Changes in the area under the precision-recall curve (aucpr), true positive rate (tp), false positive rate (fp), true-skill statistics (tss), false negative rate (fn), and true negative rate (tn) with an increase in the number of interactions used to trained the predictive models. The top six panels used 1% of the European Metaweb as validation dataset, and the bottom six panels used the interactions from Globi as validation dataset. In red are the models trained using the European Metaweb, and in blue models trained using interactions extracted from Globi. For tp, tss, fn, fp, and tn, we used the threshold that maximized the tss to transform continuous predictions into binary predictions.*

## Prediction all interactions



*Figure S7.2: Predictability of all interactions (prey and predators) of each species. Species are grouped by order on the y axis. The performance metrics are: area under operating receiver curve (auc), true-skill statistics (tss), true positive rate (tp), false positive rate (fp), true negative rate (tn), and false negative rate (fn). For tp, tss, fn, fp, and tn, we used the threshold that maximized the tss to transform continuous predictions into binary predictions.*

## Prey prediction

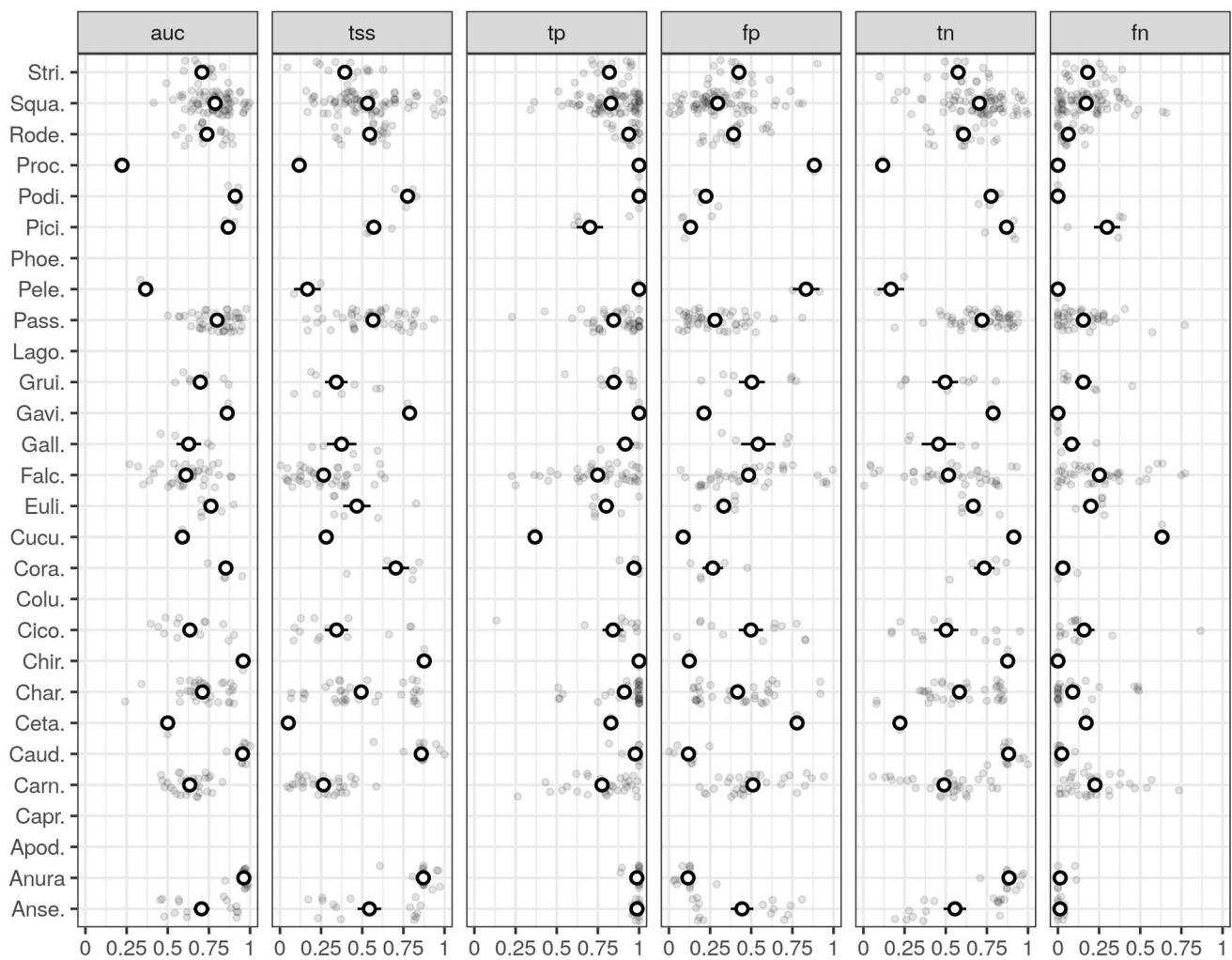
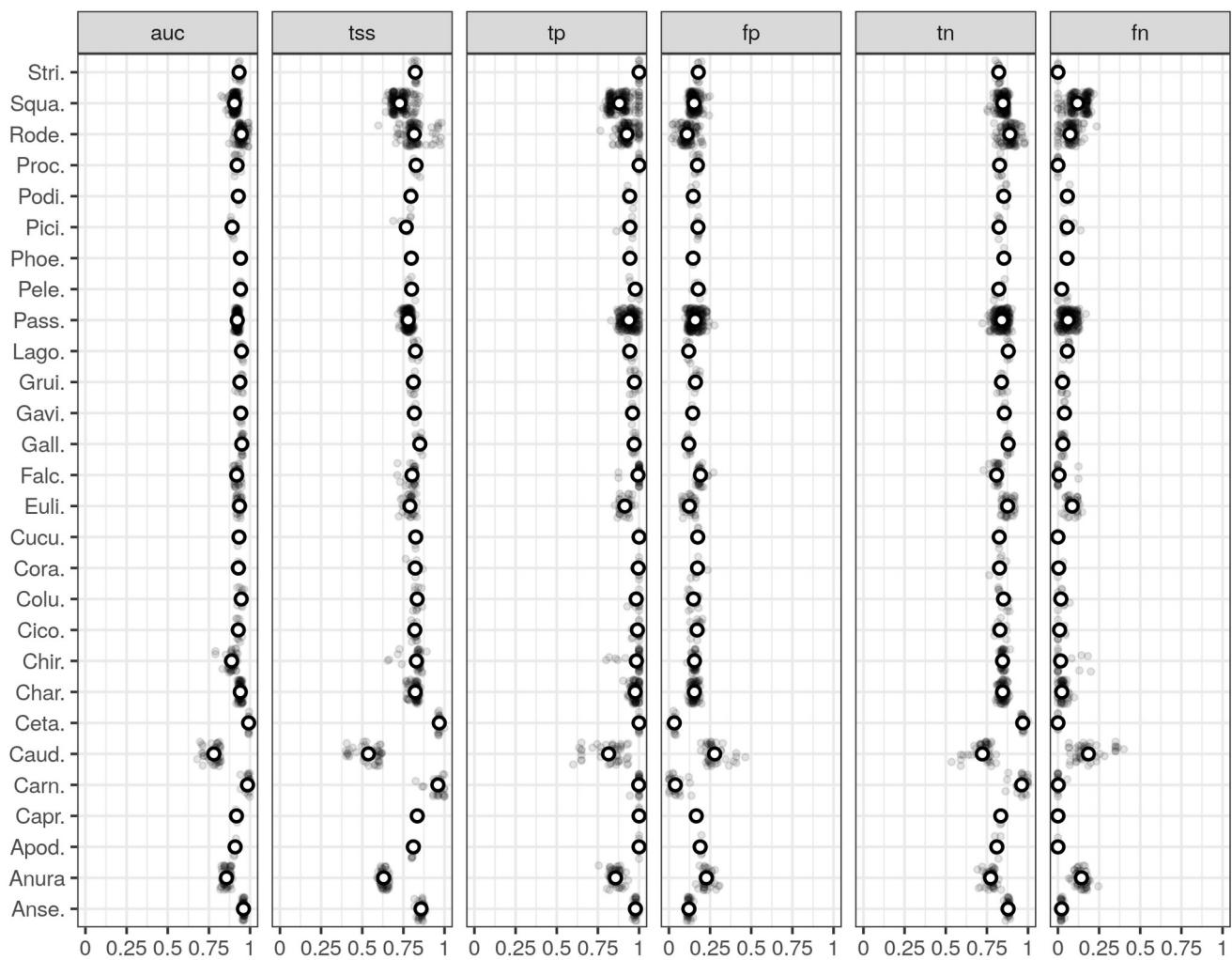


Figure S7.3: Predictability of the prey of each species. Species are grouped by order on the y axis. The performance metrics are: area under operating receiver curve (auc), true-skill statistics (tss), true positive rate (tp), false positive rate (fp), true negative rate (tn), and false negative rate (fn). For tp, tss, fn, and tn, we used the threshold that maximized the tss to transform continuous predictions into binary predictions.

## Predator prediction



*Figure S7.4: Predictability of the predators of each species. Species are grouped by order on the y axis. The performance metrics are: area under operating receiver curve (auc), true-skill statistics (tss), true positive rate (tp), false positive rate (fp), true negative rate (tn), and false negative rate (fn). For tp, tss, fn, and tn, we used the threshold that maximized the tss to transform continuous predictions into binary predictions.*

## B. Supplementary material for chapter 2

### **Trophic interaction models predict interactions across regions, not food web properties.**

Dominique Caron, Ulrich Brose, Miguel Lurgi, F. Guillaume Blanchet, Dominique Gravel, Laura J. Pollock

#### **Table of Content:**

**Appendix S1 – Description of the food webs:** Brief description of how the four food web were collated, and how we extracted predator-prey interactions.

**Appendix S2 – Trait coverage and imputation:** Additional information on the coverage of the trait data and on the fit of the MissForest algorithm used for data imputation

**Appendix S3 – Description of the structure of the predictive model:** We detail the structure of the predictive model (response variable, predictors, transformations, priors).

**Appendix S4 - Description of species' roles and food web-level properties:** We describe each species role and food web-level metrics, how we calculated them, and their relation to functioning.

**Appendix S5 - Boosted regression tree methods and results:** Brief description of the method used to fit and test the boosted regression trees and the associated main results.

**Appendix S6 - Convergence of the predictive models:** Distribution of rank-normalized potential scale reduction factor on split chains (Rhat) and rank plots of the population-level effects.

**Appendix S7 - Predicting species interactions extended results:** Results of the predicting species interactions with other evaluation metrics, and results of the linear models between model performance and distances estimates.

**Appendix S8 - Predicting species roles extended results:** Additional results for the prediction of species' role (linear model R<sup>2</sup> and coefficients).

**Appendix S9 - Predicting food web properties extended results:** Additional results for the prediction of food web results.

**Appendix S10 – References:** References for the supplemental material

## **Appendix S1: Description of the food webs**

*Europe food web:* The food web comprising European tetrapods was assembled by Maiorano et al. (2020). It contains the potential trophic interactions between all terrestrial vertebrates in Europe based on data from atlases, field guides, published papers, and expert knowledge. Maiorano et al. (2020) originally reported interactions between predators and their prey at different life-stage (eggs, larvae/young, or adult). In this study, we focused on interactions between predators and adult prey only. We combined the layer of the continental European bioregions (version 3) European Environment Agency (<https://www.eea.europa.eu/data-and-maps/data/biogeographical-regions-europe-3>) to define its spatial domain, which covers an area of  $11 \times 10^6 \text{ km}^2$ .

*Pyrenees food web:* The Pyrenees vertebrate food web was assembled independently from the European food web by Lurgi et al. (2012) who compiled all known interactions between vertebrate species in the alpine ecosystem of the southeastern Pyrenees mountain range. Data was collated from atlases, existing databases, published scientific articles and reports, and expert knowledge. This network is described at the species level and for this study, we excluded fish species. We defined the spatial domain of the food web to be of the European Mountain Area layer provided by the European Environment Agency (<https://www.eea.europa.eu/data-and-maps/data/european-mountain-areas>), covering an area of  $50 \times 10^3 \text{ km}^2$ .

*Northern Québec and Labrador food web:* The trophic tundra food web of Labrador and northern Quebec (Canada) was compiled by Berteaux et al. (2018). The Northern Québec and Labrador food web documents the presence (absence) of trophic relationships between all birds, mammals, and groups of plants (e.g., moss, ferns, grass) and arthropods (e.g., Lepidoptera, Diptera, Arachnida) whose distribution overlap Labrador and Quebec north of  $50^\circ\text{N}$  (Nunavik region). Trophic interactions were collated from a literature review, data from existing databases, and expert knowledge. For our study, we focused on the bird and mammal species (i.e., the vertebrates) of this network. We clipped at Latitude  $50^\circ\text{N}$  the polygon of the Québec province and removed Newfoundland from the Newfoundland and Labrador province layer provided by the Open Canada data portal (<https://open.canada.ca/data/en/dataset/a883eb14-0c0e-45c4-b8c4-b54c4a819edb>) to define its spatial domain, which covers  $1.3 \times 10^6 \text{ km}^2$ .

*Serengeti food web:* The Serengeti food web compiles all interactions between resource categories (e.g., seeds, shrubs, fruits and nectar), invertebrate groups (e.g., thrips, bees, moths, crickets), and vertebrate species in the Serengeti National Park (Tanzania) and was assembled from a literature review (de Visser et al., 2011). The nodes of the original food webs are resource categories, or trophic groups including one or more invertebrate groups and vertebrate species. In this study, we focused on vertebrate species, and assumed that species within a trophic group share the same predator and prey species. We used the Serengeti ecosystem layer provided by Serengeti GIS & Data Center (<https://serengetidata.weebly.com/boundaries.html>) to define its spatial domain, which covers  $36 \times 10^3$  km $^2$ .

## Appendix S2: Trait coverage and imputation

We imputed missing trait data using the *missForest* function from the *missForest* package in R (Stekhoven, 2013; Stekhoven & Buehlmann, 2012).

For each variable with missing values, the function fit a random forest using the observed values to predict missing values. The imputation procedure is repeated until a stopping criterion is met. Like for other bootstrap aggregating (bagging) algorithm, the mean of the predictions made by the individual trees of the random forest is used for continuous variables and the mode for categorical variables.

We ran the function for birds, mammals, amphibians, and reptiles separately. Each time, we included species that occur at least one of the food web (we ran *missForest* for all birds, all amphibians, all mammals, and all reptiles across the four food webs). We used all traits available (see following tables), including trait not used for interaction predictions, and the order, family, and genus of the species in the *missForest* function. This approach allows to (1) make sure a species present at different location as the same trait values, (2) take into account the large functional differences across class, (3) maximize the available data for the trait imputation within class. Not all traits were included in the predictive models to reduce collinearity among predictors (e.g., body length and body mass; age of maturity and longevity). The traits used in predictive models are bolded in the following tables.

Here, we document the trait coverage and imputation performance for each class. The traits in bold are the one used for interaction prediction (the other traits are not used in the following analyses). The performance metric is the normalized root squared mean squared error (NRMSE) for the continuous missing values, and proportion of falsely classified (PFC) for the categorical missing values. For both performance metrics, the closer to 0, the better the fit. The performance metrics are defined as follow:

$$NRMSE = \sqrt{\frac{mean((X_{true} - X_{imp})^2)}{var(X_{true})}}$$

where  $X_{true}$  is the complete continuous trait matrix,  $X_{imp}$  is the imputed continuous trait matrix. PFC is the proportion of falsely classified categorical traits.

*Amphibians (118 species):*

Trait	Variable type	Coverage (%)	Mean (SD)
<b>Trophic level (3 categories)</b>	<b>Categorical</b>	<b>62.7</b>	NA
<b>Diel Activity (2 categories)</b>	<b>Categorical</b>	<b>68.6</b>	NA
<b>Forest (Binary)</b>	<b>Categorical</b>	<b>88.1</b>	NA
<b>Savanna (Binary)</b>	<b>Categorical</b>	<b>88.1</b>	NA
<b>Shrubland (Binary)</b>	<b>Categorical</b>	<b>88.1</b>	NA
<b>Grassland (Binary)</b>	<b>Categorical</b>	<b>88.1</b>	NA
<b>Wetland (Binary)</b>	<b>Categorical</b>	<b>88.1</b>	NA
<b>Rocky areas (Binary)</b>	<b>Categorical</b>	<b>88.1</b>	NA
<b>Cave and subterranean (Binary)</b>	<b>Categorical</b>	<b>88.1</b>	NA
<b>Desert (Binary)</b>	<b>Categorical</b>	<b>88.1</b>	NA
<b>Marine (Binary)</b>	<b>Categorical</b>	<b>88.1</b>	NA
<b>Marine/Intertidal or coastal/supratidal (Binary)</b>	<b>Categorical</b>	<b>88.1</b>	NA
<b>Artificial (Binary)</b>	<b>Categorical</b>	<b>88.1</b>	NA
<b>Introduced vegetation (Binary)</b>	<b>Categorical</b>	<b>88.1</b>	NA
<b>Habitat breadth</b>	<b>Continuous</b>	<b>88.1</b>	<b>10.92 (7.92)</b>
Body length (mm)	Continuous	91.5	103.11 (56.2)
<b>Body mass (g)</b>	<b>Continuous</b>	<b>55.1</b>	<b>44.1 (174)</b>
Age of Maturity (day)	Continuous	59.3	876.6 (396)
<b>Maximum longevity (day)</b>	<b>Continuous</b>	<b>45.8</b>	<b>5442 (2848)</b>
<b>Clutch size (number of offspring)</b>	<b>Continuous</b>	<b>77.1</b>	<b>1980 (3899)</b>
Adult SVL (cm)	Continuous	0*	NA
Mean longevity (day)	Continuous	0*	NA
Generation length (day)	Continuous	0*	NA

\*When a trait is missing for all species, the trait is not used in the missForest algorithm

Error rate:

NRMSE: 0.64

PFC: 0.08

*Birds (714 species):*

Trait	Variable type	Coverage (%)	Mean (SD)
<b>Trophic level (3 categories)</b>	<b>Categorical</b>	<b>81.7</b>	NA
<b>Diel Activity (2 categories)</b>	<b>Categorical</b>	<b>95.6</b>	NA
<b>Forest (Binary)</b>	<b>Categorical</b>	<b>97.8</b>	NA
<b>Savanna (Binary)</b>	<b>Categorical</b>	<b>97.8</b>	NA
<b>Shrubland (Binary)</b>	<b>Categorical</b>	<b>97.8</b>	NA
<b>Grassland (Binary)</b>	<b>Categorical</b>	<b>97.8</b>	NA
<b>Wetland (Binary)</b>	<b>Categorical</b>	<b>97.8</b>	NA
<b>Rocky areas (Binary)</b>	<b>Categorical</b>	<b>97.8</b>	NA
<b>Cave and subterranean (Binary)</b>	<b>Categorical</b>	<b>97.8</b>	NA
<b>Desert (Binary)</b>	<b>Categorical</b>	<b>97.8</b>	NA
<b>Marine (Binary)</b>	<b>Categorical</b>	<b>97.8</b>	NA
<b>Marine/Intertidal or coastal/supratidal (Binary)</b>	<b>Categorical</b>	<b>97.8</b>	NA
<b>Artificial (Binary)</b>	<b>Categorical</b>	<b>97.8</b>	NA
<b>Introduced vegetation (Binary)</b>	<b>Categorical</b>	<b>97.8</b>	NA
<b>Habitat breadth</b>	<b>Continuous</b>	<b>97.8</b>	<b>12.75 (11.6)</b>
<b>Body mass (g)</b>	<b>Continuous</b>	<b>99.2</b>	<b>715 (4343)</b>
Age of Maturity (day)	Continuous	67.8	645 (448)
<b>Maximum longevity (day)</b>	<b>Continuous</b>	<b>70.2</b>	<b>6871 (4602)</b>
<b>Clutch size (number of offspring)</b>	<b>Continuous</b>	<b>98.6</b>	<b>4.33 (2.12)</b>
Adult SVL (cm)	Continuous	30.1	48.2 (24.7)
Mean longevity (day)	Continuous	70.2	5697 (3433)
Generation length (day)	Continuous	98.6	2466 (1515)

Error rate:

NRMSE: 0.49

PFC: 0.09

*Mammals (450 species):*

Trait	Variable type	Coverage (%)	Mean (SD)
<b>Trophic level (3 categories)</b>	<b>Categorical</b>	<b>95.8</b>	NA
<b>Diel Activity (2 categories)</b>	<b>Categorical</b>	<b>94.9</b>	NA
<b>Forest (Binary)</b>	<b>Categorical</b>	<b>91.6</b>	NA
<b>Savanna (Binary)</b>	<b>Categorical</b>	<b>91.6</b>	NA
<b>Shrubland (Binary)</b>	<b>Categorical</b>	<b>91.6</b>	NA
<b>Grassland (Binary)</b>	<b>Categorical</b>	<b>91.6</b>	NA
<b>Wetland (Binary)</b>	<b>Categorical</b>	<b>91.6</b>	NA
<b>Rocky areas (Binary)</b>	<b>Categorical</b>	<b>91.6</b>	NA
<b>Cave and subterranean (Binary)</b>	<b>Categorical</b>	<b>91.6</b>	NA
<b>Desert (Binary)</b>	<b>Categorical</b>	<b>91.6</b>	NA
<b>Marine (Binary)</b>	<b>Categorical</b>	<b>91.6</b>	NA
<b>Marine/Intertidal or coastal/supratidal (Binary)</b>	<b>Categorical</b>	<b>91.6</b>	NA
<b>Artificial (Binary)</b>	<b>Categorical</b>	<b>91.6</b>	NA
<b>Introduced vegetation (Binary)</b>	<b>Categorical</b>	<b>91.6</b>	NA
<b>Habitat breadth</b>	<b>Continuous</b>	<b>91.6</b>	<b>5.87 (5.32)</b>
<b>Body mass (g)</b>	<b>Continuous</b>	<b>96.4</b>	<b>34958 (235221)</b>
Age of Maturity (day)	Continuous	63.8	460 (563)
<b>Maximum longevity (day)</b>	<b>Continuous</b>	<b>75.3</b>	<b>4784 (4440)</b>
<b>Clutch size (number of offspring)</b>	<b>Continuous</b>	<b>86.2</b>	<b>3.33 (2.00)</b>
Adult SVL (cm)	Continuous	80.0	52.3 (78.3)
Mean longevity (day)	Continuous	75.3	4410 (4237)
Generation length (day)	Continuous	95.6	1535 (1154)

Error rate:

NRMSE: 0.71

PFC: 0.12

*Reptiles (283 species):*

Trait	Variable type	Coverage (%)	Mean (SD)
<b>Trophic level (3 categories)</b>	<b>Categorical</b>	<b>68.2</b>	NA
<b>Diel Activity (2 categories)</b>	<b>Categorical</b>	<b>76.0</b>	NA
<b>Forest (Binary)</b>	<b>Categorical</b>	<b>64.0</b>	NA
<b>Savanna (Binary)</b>	<b>Categorical</b>	<b>64.0</b>	NA
<b>Shrubland (Binary)</b>	<b>Categorical</b>	<b>64.0</b>	NA
<b>Grassland (Binary)</b>	<b>Categorical</b>	<b>64.0</b>	NA
<b>Wetland (Binary)</b>	<b>Categorical</b>	<b>64.0</b>	NA
<b>Rocky areas (Binary)</b>	<b>Categorical</b>	<b>64.0</b>	NA
<b>Cave and subterranean (Binary)</b>	<b>Categorical</b>	<b>64.0</b>	NA
<b>Desert (Binary)</b>	<b>Categorical</b>	<b>64.0</b>	NA
<b>Marine (Binary)</b>	<b>Categorical</b>	<b>64.0</b>	NA
<b>Marine/Intertidal or coastal/supratidal (Binary)</b>	<b>Categorical</b>	<b>64.0</b>	NA
<b>Artificial (Binary)</b>	<b>Categorical</b>	<b>64.0</b>	NA
<b>Introduced vegetation (Binary)</b>	<b>Categorical</b>	<b>64.0</b>	NA
<b>Habitat breadth</b>	<b>Continuous</b>	<b>64.0</b>	<b>5.29 (2.98)</b>
<b>Body mass (g)</b>	<b>Continuous</b>	<b>95.8</b>	<b>2559 (29423)</b>
Age of Maturity (day)	Continuous	42.4	975 (990)
<b>Maximum longevity (day)</b>	<b>Continuous</b>	<b>64.0</b>	<b>5350 (6840)</b>
<b>Clutch size (number of offspring)</b>	<b>Continuous</b>	<b>83.4</b>	<b>7.35 (8.68)</b>
Adult SVL (cm)	Continuous	44.5	10.16 (12.92)
Mean longevity (day)	Continuous	40.6	4128 (5070)
Generation length (day)	Continuous	0*	NA

\*When a trait is missing for all species, the trait is not used in the missForest algorithm

Error rate:

NRMSE: 0.78

PFC: 0.11

## **Appendix S3: Description of the structure of the predictive model**

The predictive models we use is a Bayesian hierarchical generalized linear model with 13 trait-based predictor variables with varying intercepts and slopes for the order of the predator. All traits were extracted from Etard et al. (2020). All traits are described in the original paper, but here, we describe briefly each derived variable used in the predictive models:

- Population-level effects:
  - Trophic level of the predator: Categorical with 3 levels - Herbivore, Omnivore, and Carnivore. Coded as 2 binary variables (Omnivore and Carnivore).
  - Habitat Breadth of the predator: Quantitative - Number of habitats the predator is known to use based on the second level of the IUCN habitat data.
  - Body mass of the predator: Quantitative - Log-transformed body mass of the predator
  - Longevity of the predator: Quantitative - Life span of an individual of the predator species. Species that live longer can be expected to be predators has they favour survival over reproduction.
  - Clutch Size of the predator: Quantitative - Number of offspring or eggs per clutch/litter by the predator. Species that have less offsprings can be expected to be predators has they favour survival over reproduction.
  - Trophic level of the prey: Categorical with 3 levels - Herbivore, Omnivore, and Carnivore. Coded as 2 binary variables (Omnivore and Carnivore)
  - Habitat Breadth of the prey: Quantitative - Number of habitats the prey is known to use based on the second level of the IUCN habitat data.
  - Body mass of the prey: Quantitative - Log-transformed body mass of the prey
  - Longevity of the prey: Quantitative - Life span of an individual of the prey species. Species that have shorter life can be expected to be prey has they favour reproduction over survival.
  - Clutch Size of the prey: Quantitative - Number of offspring or eggs per clutch/litter by the prey. Species that have more offsprings can be expected to be prey has they favour reproduction over survival.
  - Activity Time match: Categorical binary – The predator and the prey are active during the same time of the day (1) or not (0). The original activity time data for the prey and the

predator is a binary variable with possible values nocturnal and non-nocturnal (diurnal, crepuscular, or cathemeral)

- Habitat match: Quantitative – Jaccard similarity between the habitat use of the predator and the prey. The original habitat data are 12 binary variables: Forest, Savanna, Shrubland, Grassland, Wetland, Rocky areas, Cave and subterranean, Desert, Marine, Marine/Intertidal or coastal/supratidal, Artificial, Introduced vegetation
- Body mass match: The squared difference between the log-transformed body mass of the predator and the log-transformed body mass of the prey.

We centered and scaled the quantitative predictors by subtracting each value by the mean and dividing by twice the standard deviation. The mean and standard deviation of each predictor was computed over the entire set of species (over all food webs). The predictors that we scaled are: the log-transformed habitat breadth of the predator and the prey, the log-transformed body mass of the predator and the prey, the log-transformed longevity of the predator and the prey, the log-transformed clutch/litter size of the predator and the prey, the habitat match, and the body mass match.

- Group-level effects (intercept and slopes):
  - Order of the predator
    - Northern Québec and Labrador food web: 21 levels
    - Europe: 36 levels
    - Pyrenees: 23 levels
    - Serengeti: 41 levels
- Model:

$$L_{ij} \sim \text{Bernouilli}(p_{ij}),$$

$$\text{logit}(p_{ij}) = \alpha + \alpha_{\text{predator}[j]} + \sum_{k=1}^{13} (\beta_k \times T_k + \beta_{k,\text{predator}[j]} \times T_k),$$

- $L_{ij}$  is the interaction between predator  $j$  and prey  $i$ .
- $P_{ij}$  is the probability of interaction between predator  $j$  and prey  $i$

- $\alpha$  and  $\beta$  are the fixed intercept and slopes
  - $\alpha_{predator[j]}$  and  $\beta_{predator[j]}$  are the order-specific intercept and slopes for the order of predator  $j$
  - $T_k$  are the trait-based predictors.
- Priors:
$$\alpha, \beta \sim Normal(0, 1)$$

$$\alpha_{predator[j]}, \beta_{k,predator[j]} \sim Normal(\mu, \sigma)$$

$$\mu \sim Normal(0, 1)$$

$$\sigma \sim HalfCauchy(0, 5)$$
- Number of iterations:
  - Warmup: 2000 iterations
  - Inference: 2000 iterations
- Training:
  - 70% of all interactions in each food web and a equal number of non-interactions (random undersampling): Northern Québec and Labrador food web (769 interactions and non-interactions), Europe (40 422 interactions and non-interactions), Pyrenees (582 interactions and non-interactions), Serengeti (7727 interactions and non-interactions). Random undersampling is a strategy to handle imbalanced dataset where one class of the response variable (non-interactions) is a lot more prevalent than the other class (interactions). Imbalanced dataset can influence and hinder training of some predictive models.
- Validation:
  - For within food web predictions (i.e., trophic interaction predictions in the food web used for training): We used the remaining 30% of the interactions of each food webs and a number of non-interactions that makes the prevalence of interactions in the validation dataset equal to the prevalence of interactions in the entire food web.
  - For across food web predictions (i.e., trophic interaction predictions in food webs other than the one used for training): We used the entire food web (all interactions and non-interactions).

- Model formula:

```
brms_form <- bf(interaction ~ 1 + (Omnivore.predator +
Carnivore.predator + Habitat_breadth.predator + BM.predator +
Longevity.predator + ClutchSize.predator + Omnivore.prey +
Carnivore.prey + Habitat_breadth.prey + BM.prey + Longevity.prey +
ClutchSize.prey + ActivityTime.match + Habitat.match + BM.match) +
(1 + (Omnivore.predator + Carnivore.predator +
Habitat_breadth.predator + BM.predator + Longevity.predator +
ClutchSize.predator + Omnivore.prey + Carnivore.prey +
Habitat_breadth.prey + BM.prey + Longevity.prey + ClutchSize.prey +
ActivityTime.match + Habitat.match + BM.match) || Order.predator),
family = bernoulli())
```

## Appendix S4: Description of species roles and food web-level properties

### Species roles:

For more information on what each species role metrics represent and aim to describe, we recommend the reviews made by Cirtwill et al. (2018) and Delmas et al. (2019). Here, we very briefly describe what these metrics are, how they are related to the functioning and the stability of food webs:

- **Number of prey:** We calculated the number of prey of a species as the number of interactions for which the species is the predator (in-degree). It measures how generalized the species is. The loss of species with many prey are more likely to trigger top-down cascades (Curtsdotter et al., 2011). In contrast, specialist predators are more vulnerable to co-extinction. In-degree and out-degree quantify the direct influence of species on other species.
- **Number of predators:** We calculated the number of predators of a species as the number of interactions for which the species is the prey (out-degree). It measures the vulnerability of the prey. The loss of species with many predators are more likely to trigger bottom-up cascades (Curtsdotter et al., 2011). In-degree and out-degree quantify the direct influence of species on other species.
- **Betweenness:** The betweenness centrality of a species is the number of times the shortest path linking any two species in the network goes through the focal species (Freeman, 1977). Betweenness quantify the number of food chains a species takes part in, and thus how it affects energy flows (Cirtwill et al., 2018). It is considered one important measures of species topological importance (Jordán, 2009). We calculated betweenness using the function *centr\_betw* in the R package *igraph* (Csardi & Nepusz, 2006).
- **Closeness:** Closeness centrality measures how far a species is to any other species. It is therefore a more “global” measure of centrality than betweenness and quantify how “efficiently” the change in this species will spread through the food web. We used the function *centr\_clo* in the R package *igraph* (Csardi & Nepusz 2006).

- **Eigenvector centrality:** Eigenvector centrality uses the eigenvector of the largest eigenvalue of the adjacency matrix describing the food web. It is another “global” measure of centrality that aims to describe species importance in the food web. Eigenvector centrality is related to the contribution of species to network stability (Allesina & Pascual, 2009). We used the function *centr\_eigen* of the R package *igraph* (Csardi & Nepusz 2006).
- **Trophic level:** Trophic level describe the “vertical” position (primary consumer, secondary consumer, etc.) of a species in a food web. We calculated trophic level using the function *TrophInd* of the package *NetIndices* package (Kones et al., 2009). The function calculate trophic level as  $1 + \text{the weighted average of the trophic levels of its food items}$ . Top predators and basal species are expected to have larger effect on the community through top-down and bottom-up cascades (Dyer & Letourneau, 2003).
- **Omnivory:** Omnivory is defined as the variety of trophic levels a species feed on. Omnivory has the potential to stabilize food webs by mitigating trophic cascades, although this effect is debated (Bascompte et al., 2005). We calculated omnivory using the function *TrophInd* of the package *NetIndices* package (Kones et al. 2009).
- **Within-module degree:** Within-module degree quantify how a species is connected within its module (Guimerà & Amaral, 2005). Modules (groups of species interacting more together than with other species) are first defined using a spin-glass model and simulated annealing using the function *cluster\_springlass* from the package *igraph* (Csardi & Nepusz 2006). We used the function *calc\_topological\_roles* of the R package *multiweb* (Saravia 2022) to calculate within-module degree. Species with high within-module degree are important in the stability and functioning of their module.
- **Participation coefficient:** Participation coefficient quantify how much a species interact with species in other module (Guimerà & Amaral 2005). A participation coefficient equal to 0 defines a species solely interacting within its module, whereas 1 defines a species interacting with species with all other modules equally. Species with high participation coefficient connects module together, and therefore, are important in the cohesion of the entire food web. We used

the function *calc\_topological\_roles* of the R package *multiweb* (Saravia, 2022) to calculate participation coefficient.

- **Motif profile:** The motif profile of a species is the frequency of which a species occupies a position within each motif (Milo et al., 2002). We used 3-species motifs, which are the 13 different arrangements 3 species can take. For 3-species motifs, there are 30 different positions a species can take. The motif profile has been proposed to be a more complete description of the Eltonian niche of a species (Cirtwill et al., 2018; Stouffer et al., 2012). Some motifs such as “linear chain”, “intraguild predation”, “direct competition”, and “apparent competition” influence the stability of food web (Borrelli, 2015; Stouffer & Bascompte, 2010).

### **Food web properties:**

- **Connectance:** Connectance is defined as the proportion of interactions relative to the total number of possible interactions. Here, since we have a directed network with self-loop allowed, it is calculated as the number of interactions ( $L$ ) divided by the squared the number of species ( $S^2$ ). Connectance relates to community sensitivity to perturbations and influence its dynamics (Dunne et al., 2002; Vieira & Almeida-Neto, 2015).
- **Mean and maximum trophic level:** The mean and maximum trophic level are calculated as the average and highest trophic levels among species within a food web. These measures “vertical” diversity, describing on how many trophic levels energy flows through in the food web. Vertical diversity interacts with horizontal diversity (within trophic level diversity) to influence community stability (Duffy et al., 2007; Zhao et al., 2019).
- **Motifs distribution:** The motif distribution is the frequency of motifs (Milo et al. 2002). Here, we focused on the 13 possible 3-species motifs. As mentioned in species role section, some motifs such as “linear chain”, “intraguild predation”, “direct competition”, and “apparent competition” influence the stability of food web (Borrelli, 2015; Stouffer & Bascompte, 2010). We used the function *motif* of the package *igraph* (Csardi & Nepusz 2006) to get the motif distribution of food webs.

- **Food web diameter:** The diameter is the longest of all shortest paths between any two species in the food web. The diameter influence how “efficiently” a perturbation spread thought the entire food web. We calculated food web diameter using the function *diameter* of the package *igraph* (Csardi & Nepusz 2006).
- **Number of clusters:** The number of clusters (i.e., modules) are the number of non-overlapping highly connected species. We determined the number of clusters using the function *cluster\_springlass* from the package *igraph* (Csardi & Nepusz 2006). Perturbations spread more efficiently within a cluster than across clusters. So, an higher number of clusters should increase resistance to perturbations.
- **Modularity:** Modularity quantify how food webs are divided into modules (groups of species interacting more together than with other species). Modularity increases food web persistence (Stouffer & Bascompte, 2011). We determined modules/clusters using the function *cluster\_springlass* from the package *igraph* (Csardi & Nepusz 2006).

## **Appendix S5: Boosted regression tree methods and results**

### ***Methods***

To make sure that the boosted regression trees (BRT) uses the same amount of information than the Bayesian hierarchical generalized linear model, we used the set of 13 trait-based predictor variables described in Appendix S3. We also added a 14<sup>th</sup> predictor, the order of the predator as predictors, as this trait is used as a group-level effects in the Bayesian model. The predictors were centered and scaled as described in Appendix S3. The training and validation sets are also the exact same as the ones to train and test the Bayesian models.

We used the packages gbm (Greenwell et al., 2019) and dismo (Hijmans et al., 2017) in R to fit the BRT on each of the training set. We set the tree complexity to 5, the learning rate to 0.01, and the bag fraction to 0.5 for each model.

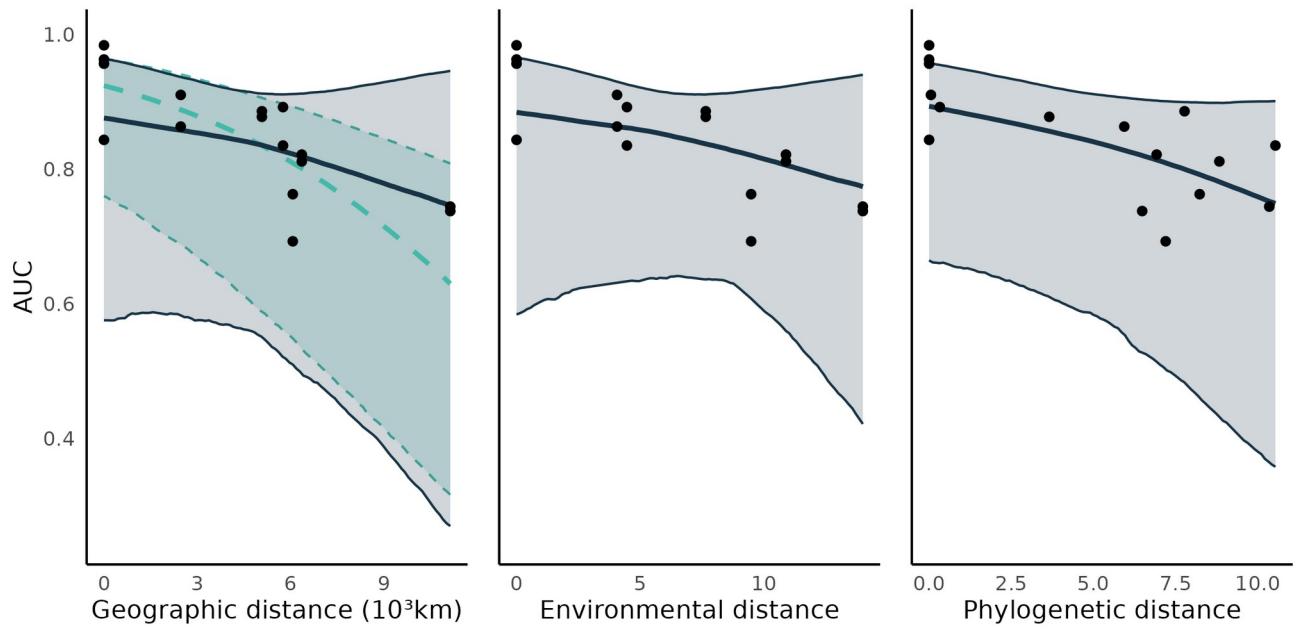
## Results

Here, we present the main results of the manuscript:

*Predicting trophic interactions:*

*Table S5.1: Area under the receiver operating curve (AUC) and area under the precision-recall-gain curve (AUPRG) of each BRT model predicting every food web. Values in parenthesis are the AUC and AUPRG obtained by the Bayesian models.*

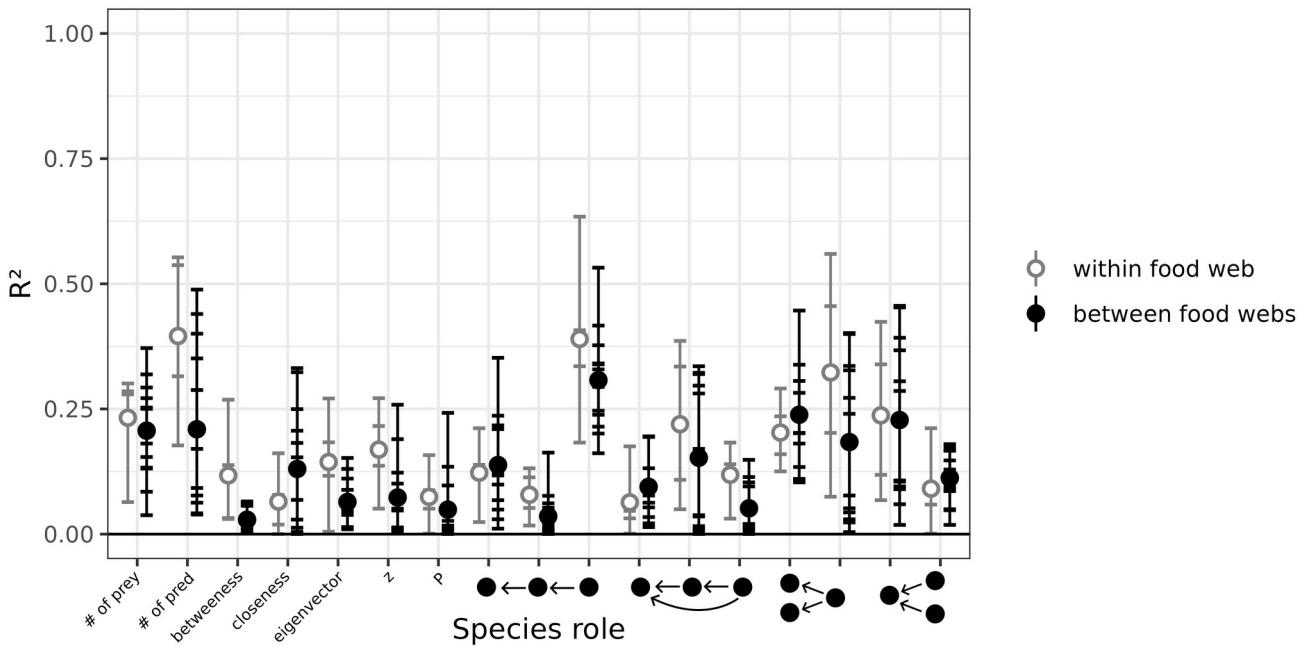
Model AUC	Europe	Pyrenees	North Québec and Labrador	Serengeti
Europe	0.98 (0.96)	0.91 (0.89)	0.89 (0.9)	0.82 (0.8)
Pyrenees	0.86 (0.86)	0.96 (0.95)	0.88 (0.85)	0.76 (0.79)
North Québec And Labrador	0.83 (0.82)	0.89 (0.9)	0.96 (0.95)	0.74 (0.75)
Serengeti	0.81 (0.85)	0.69 (0.78)	0.73 (0.77)	0.84 (0.92)
AUPRG				
Europe	0.99 (0.96)	0.81 (0.63)	0.91 (0.91)	0.80 (0.84)
Pyrenees	0.83 (0.9)	0.97 (0.95)	0.88 (0.94)	0.63 (0.84)
North Québec And Labrador	0.87 (0.88)	0.93 (0.9)	0.98 (0.93)	0.68 (0.8)
Serengeti	0.73 (0.84)	0.64 (0.73)	0.73 (0.77)	0.85 (0.93)



**Figure S5.1: Transferability of BRT models.** Each point is the predictive performance (AUC) of a BRT-food web predicted combination (16 combinations; Table S5.1). The trend lines are the median effects with their 95% credible interval constructed with the posterior predictive distribution of geographic, environmental, and phylogenetic distances on predictive performance. Shown are the total (turquoise; -0.63, CrI 95% = [-0.28, -0.96]) and partial (dark blue; -0.28, CrI 95% = [-1.12, 0.59]) effects of geographic distance , and the direct (controlling for geographic distance) effects of environmental (-0.27, CrI 95% = [-1.12, 0.59]) and phylogenetic distances (-0.41, CrI 95% = [-1.07, 0.20]).

#### Predicting species trophic role:

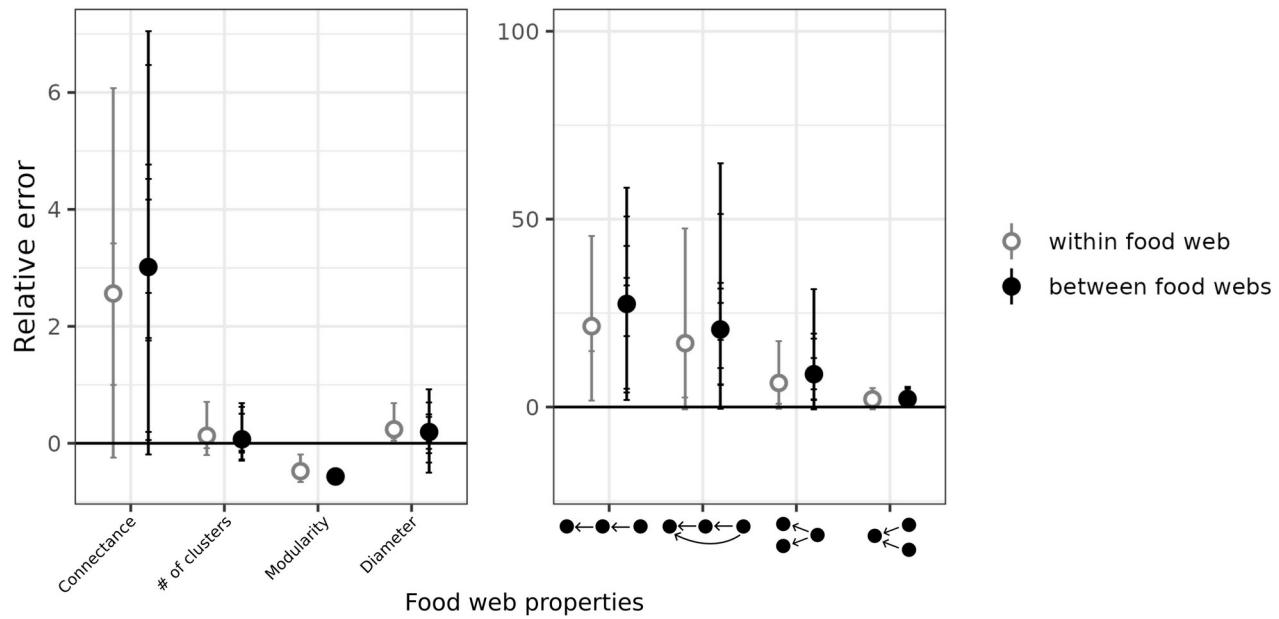
Similar to the Bayesian hierarchical models, BRTs failed to accurately predict species trophic roles (Figure S5.2, compared to Figure 4). The number of prey, within-module degree, participation coefficient, and the frequency of most motif positions were slightly better predicted by the bayesian hierarchical models, especially for within food web predictions. The remaining trophic roles were similarly well predicted by the BRTs and the Bayesian hierarchical models.



*Figure S5.2: Performance of boosted regression tree models to predict species' role measured as the proportion of the variance in trophic positions explained by the models ( $R^2$ ). From left to right, the figure shows the number of prey, number of predators, betweenness, closeness, eigenvector centrality, within-module degree ( $z$ ), participation coefficient ( $P$ ), times the species is at a specific position in the linear chain (predator – consumer - resource), intraguild predation (omnivore – consumer - resource), direct competition (2 consumers – 1 resource), and apparent competition motifs (1 consumer – 2 resources). The short horizontal lines highlight the  $R^2$  for each role, model, and food web predicted combination. Grey open dots are the mean  $R^2$  for within food web predictions. Full black dots are the mean  $R^2$  for between food web predictions.*

#### *Predicting food web properties:*

Overall, some food web properties are better predicted by the boosted regression trees than by the Bayesian models (Figure S5.3, compared to Figure 5). Connectance is especially much less overpredicted (relative error of 0-6 instead of 1-10), and the frequency of intraguild predation motifs is also much better predicted. However, we failed calculate mean and maximum trophic levels. This is likely due to the prediction of many trophic cycles (e.g., two species feeding on each other), which can create problem when defining trophic levels. The other properties predicted by the BRT are similarly biased to the properties predicted by the Bayesian models.



*Figure S5.3: Relative error of global food web properties predicted by the boosted regression trees. From left to right, the figure shows the relative error for connectance, number of clusters, modularity, diameter, number of linear chain (predator – consumer - resource), intraguild predation (omnivore – consumer - resource), direct competition (2 consumers – 1 resource), and apparent competition (1 consumer – 2 resources) motifs . Short horizontal lines are the relative errors for each property, model, and food web predicted combination. Grey open dots are the mean relative errors for within food web predictions and full black dots are the mean relative errors for between food web predictions.*

## Appendix S6: Convergence of the predictive models

We evaluated convergence and adequacy of the four predictive models using rank plots, posterior predictive checks, and the rank-normalized potential scale reduction factor on split chains.

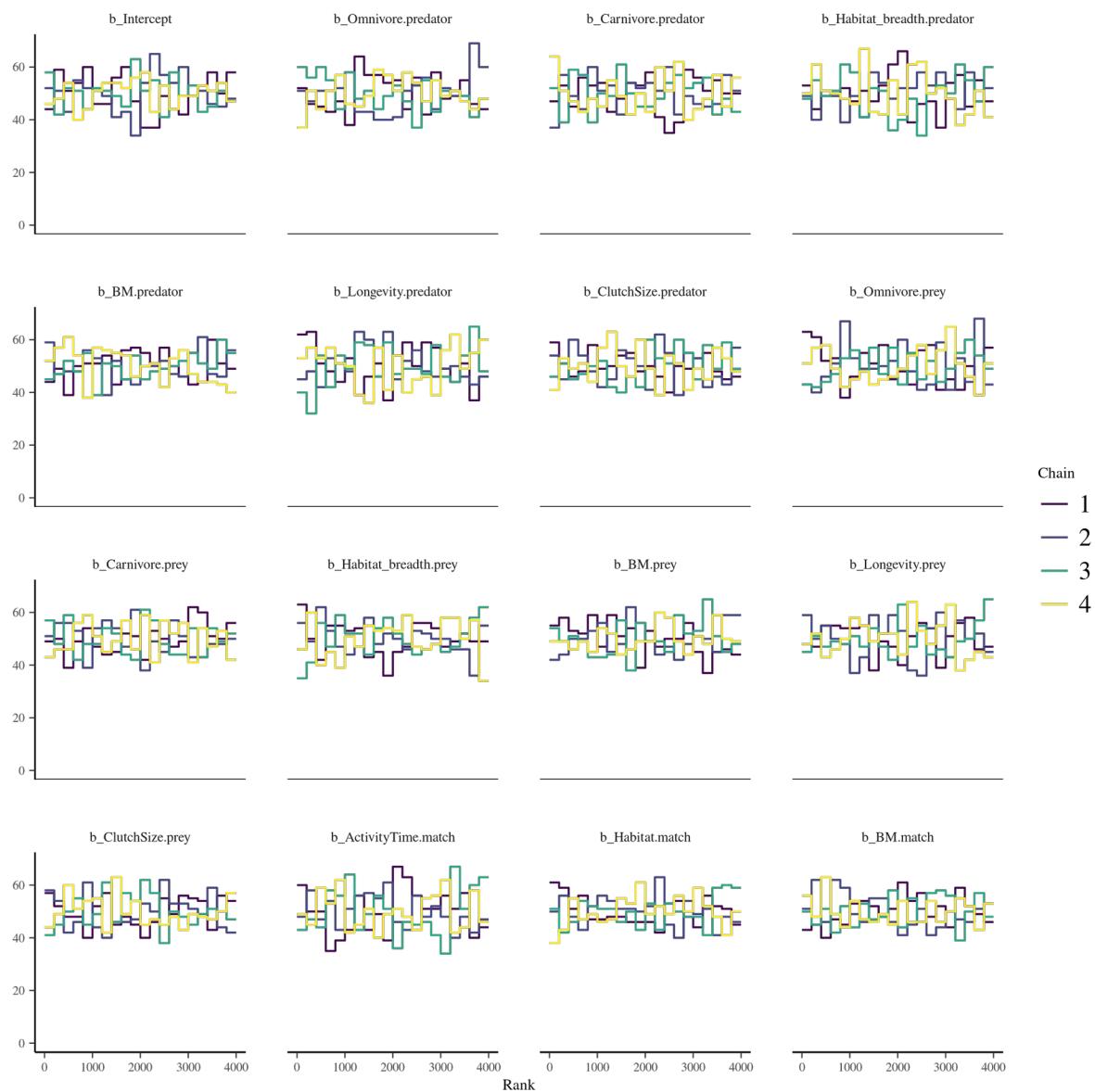
Rank trace plots are similar to trace plots and are used to assess the mixing of the chains (Gelman et al., 2013). With trace rank plots, we check if the MCMC sampler explores the parameter space adequately. For each iterations (x axis), the parameter varies (y axis), and the chains are ranked from 1 (lowest value of the parameter) to 4 (highest value of the parameter). In the ideal case, we want that the chains to change rank more or less randomly (no chain staying at a given rank more than others). Conversely, if a given chain stays in the same rank for many iterations in a row, or is more often at the same rank than other chains, it is sign of a problem with mixing. Here, all the rank plots suggest adequate mixing of the chains (Figure S6.1, S6.4, S6.7, S6.10).

Posterior predictive checks compare the distribution of the observed data to the distribution of the posterior predictive distribution (Gelman et al., 2013). A discrepancy in distributions can be caused by a model missfit. A model that fits correctly should replicate the observed distribution of interactions/non interactions (counts on the y axis, and 0 and 1 on the x axis) of each “group” of the model (in our models, we have varying parameters for the order of the predator, so an histogram for each predator order). A problematic model would fail to replicate the distributions. Here, the four models replicate the distributions of the observed data (Figure S6.2, S6.5, S6.8, S6.11).

Finally, the rank-normalized potential scale reduction factor on split chains (Rhat) is similar to the potential scale reduction factor, and is a summary statistics evaluating model convergence (Vehtari et al., 2021). Rhat close to 1 indicate a correctly converged model, and a Rhat larger than 1 ( $Rhat > 1.1$ ) would indicate that the model did not converge correctly. Here, we have a large number of Rhat (one for each slopes and intercept of each predator order), so we are plotting the histogram of Rhat. Each histogram are suggesting the adequate convergence of the models (Figure S6.3, S6.6, S6.9, S6.12).

*Northern Québec and Labrador model:*

- Trace rank plots of the population-level effects:



*Figure S6.1: Trace rank plots of the population-level effects for the model calibrated on the Northern Québec and Labrador food web.*

- Posterior predictive checks:

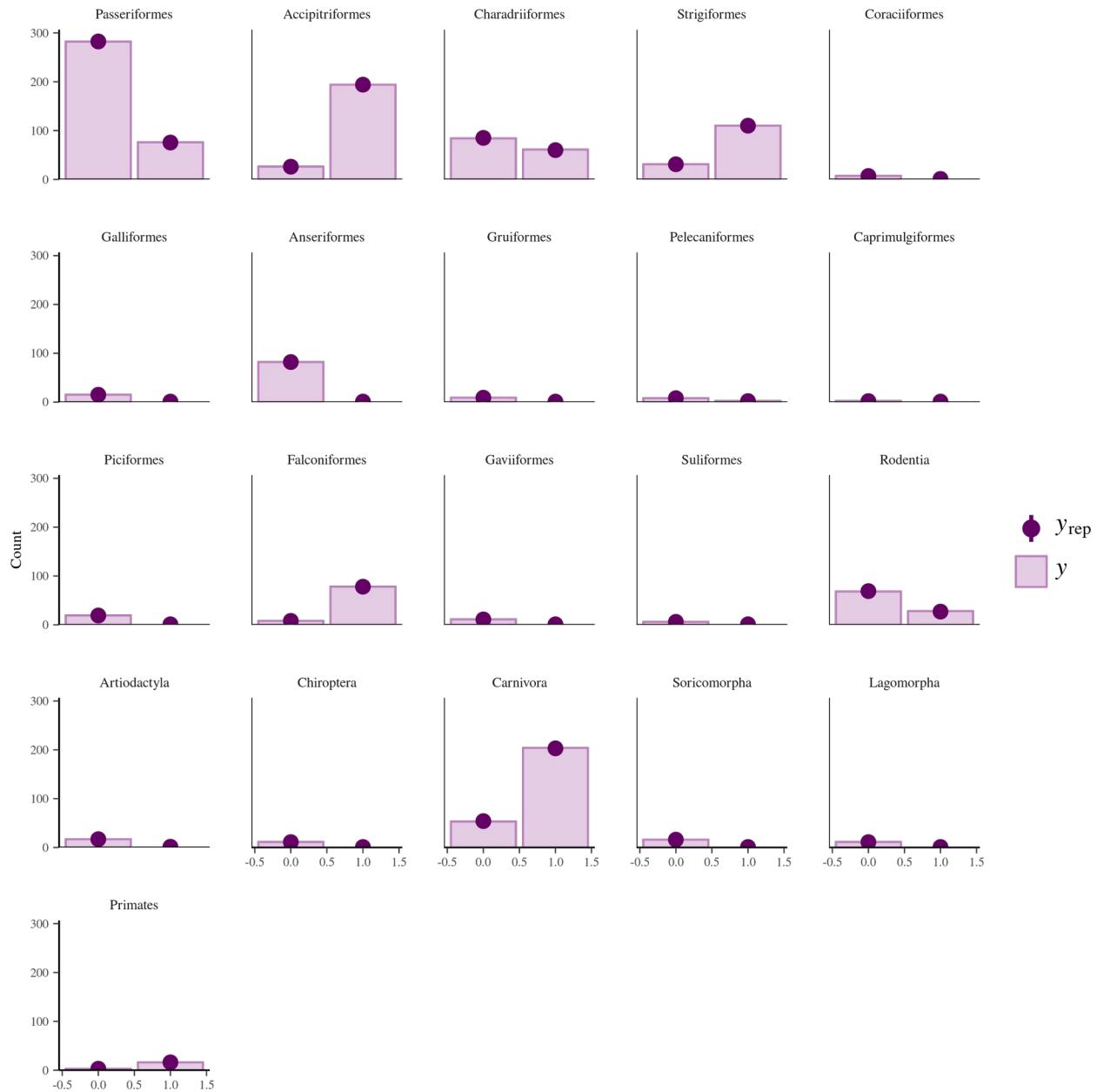
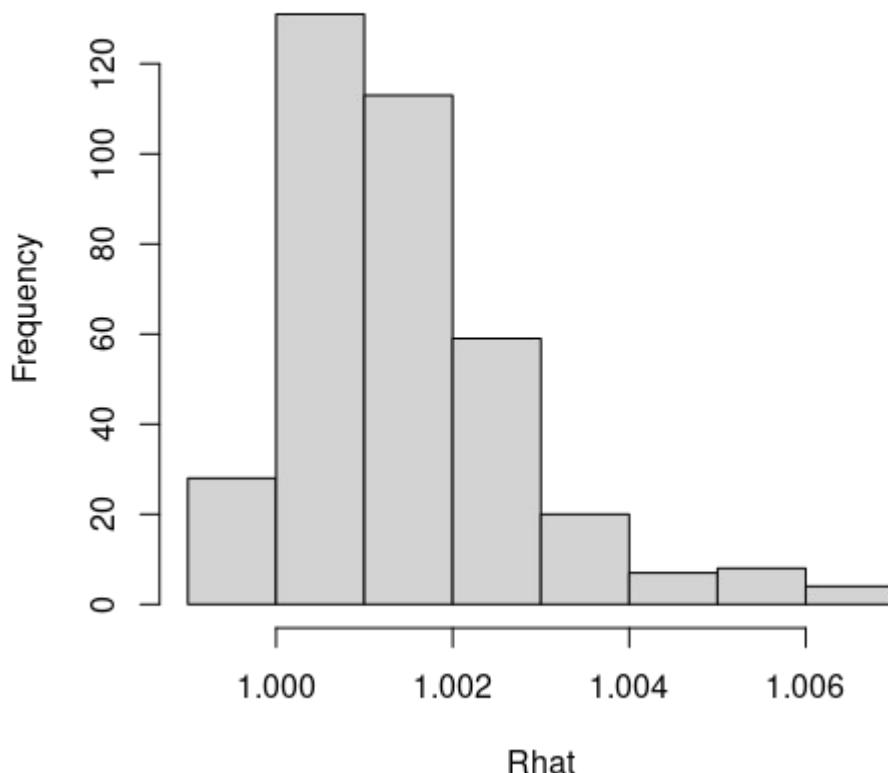


Figure S6.2: Comparison of the posterior predictive distribution ( $y_{rep}$ ; purple points) and the observed distribution ( $y$ ; purple bars) of the model calibrated on the Northern Québec and Labrador food web.

- Rank-normalized potential scale reduction factors on split chains:

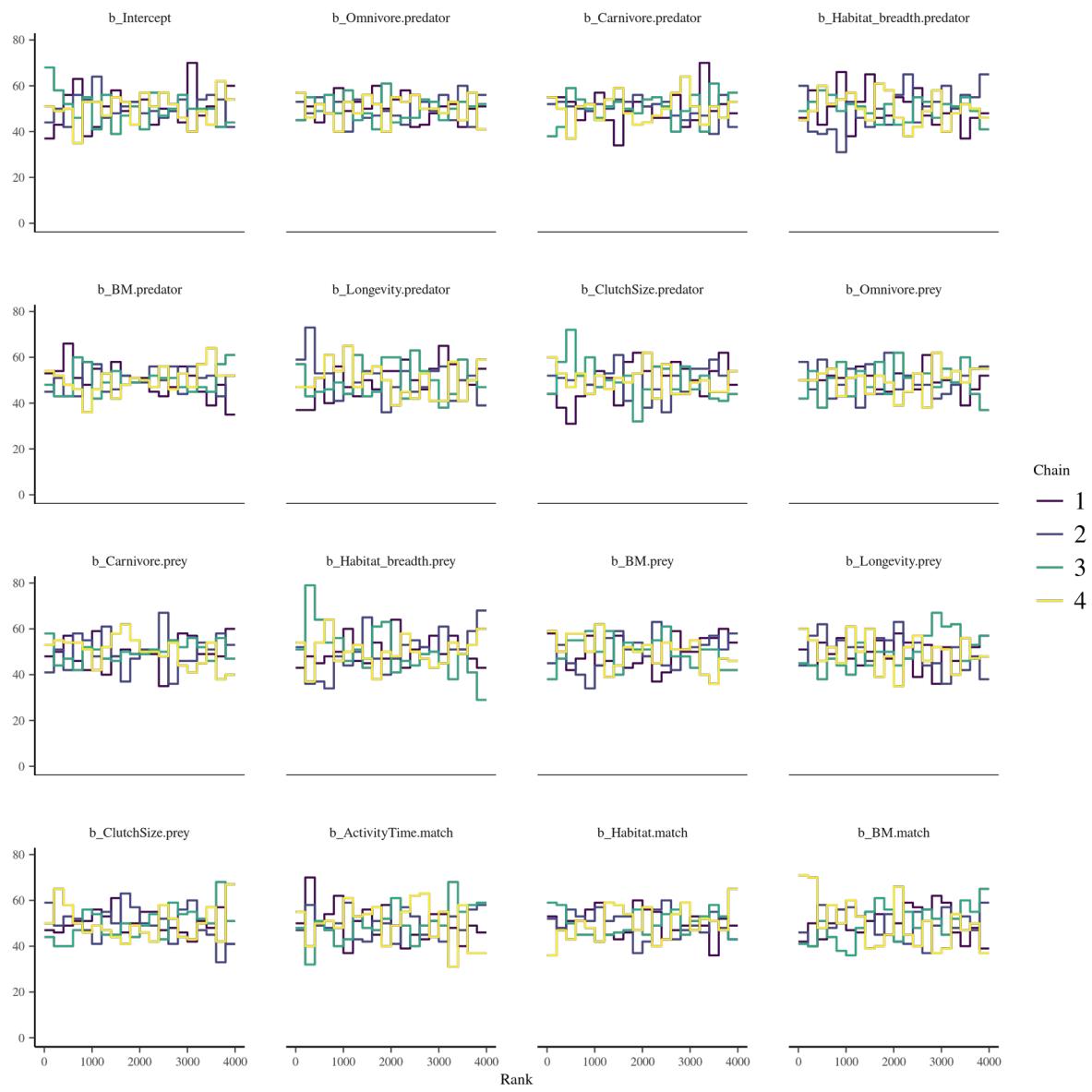
**Distribution of rank-normalized potential scale reduction factors for the Northern Québec and Labrador model**



*Figure S6.3: Distribution of the rank-normalized potential scale reduction factors on split chains for the Northern Québec and Labrador food web*

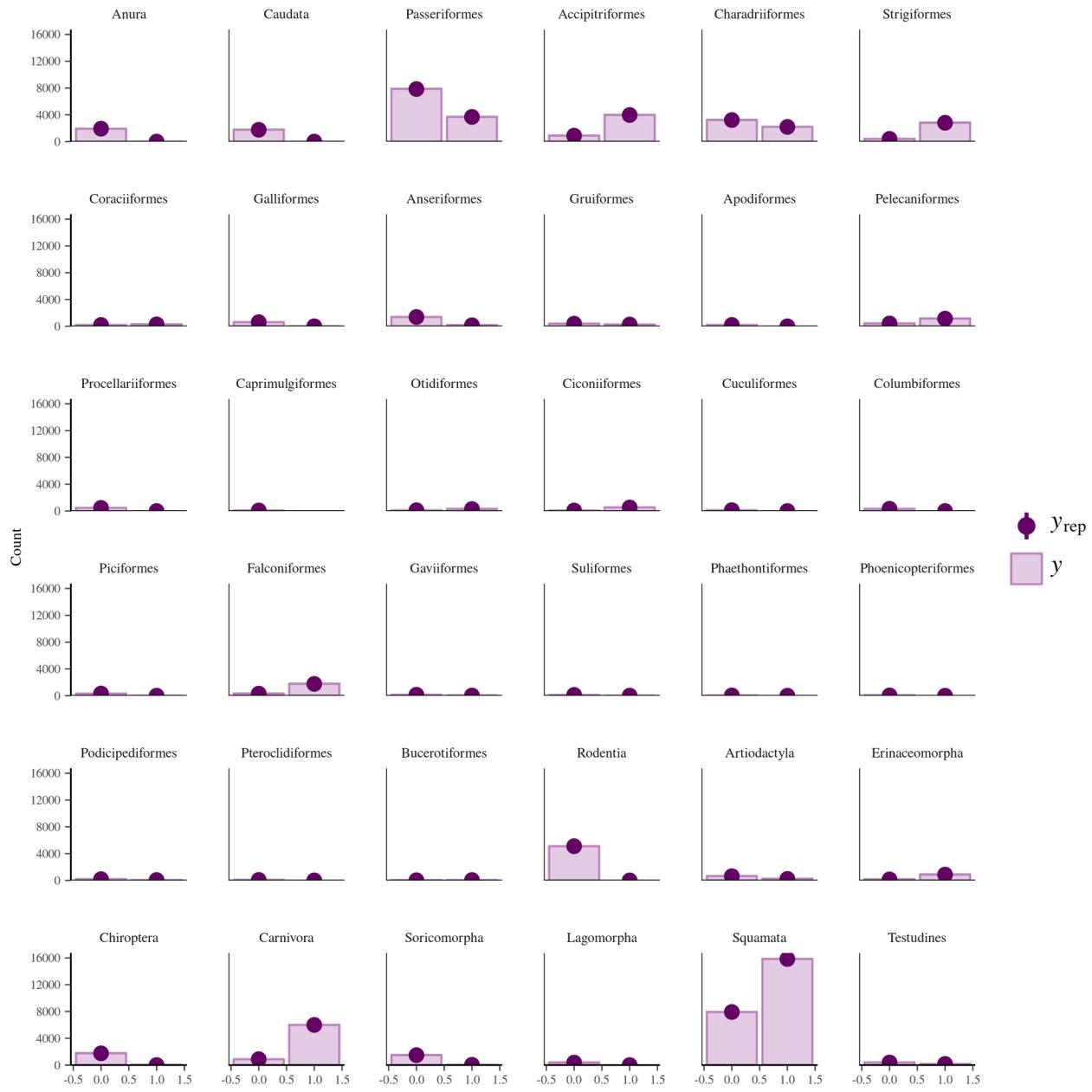
*Europe model:*

- Trace rank plots of the population-level effects:



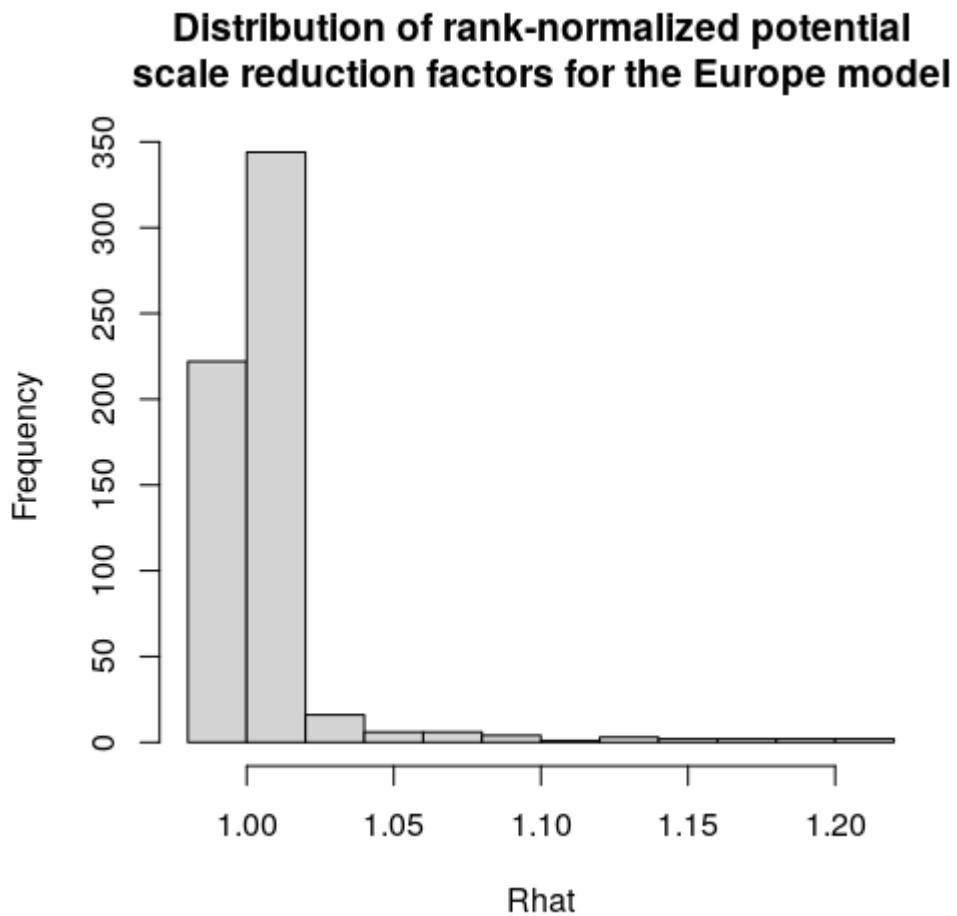
*Figure S6.4: Trace rank plots of the population-level effects for the model calibrated on the European food web of terrestrial vertebrates.*

- Posterior predictive checks:



*Figure S6.5: Comparison of the posterior predictive distribution ( $y_{rep}$ ; purple points) and the observed distribution ( $y$ ; purple bars) of the model calibrated on the food web of European terrestrial vertebrates.*

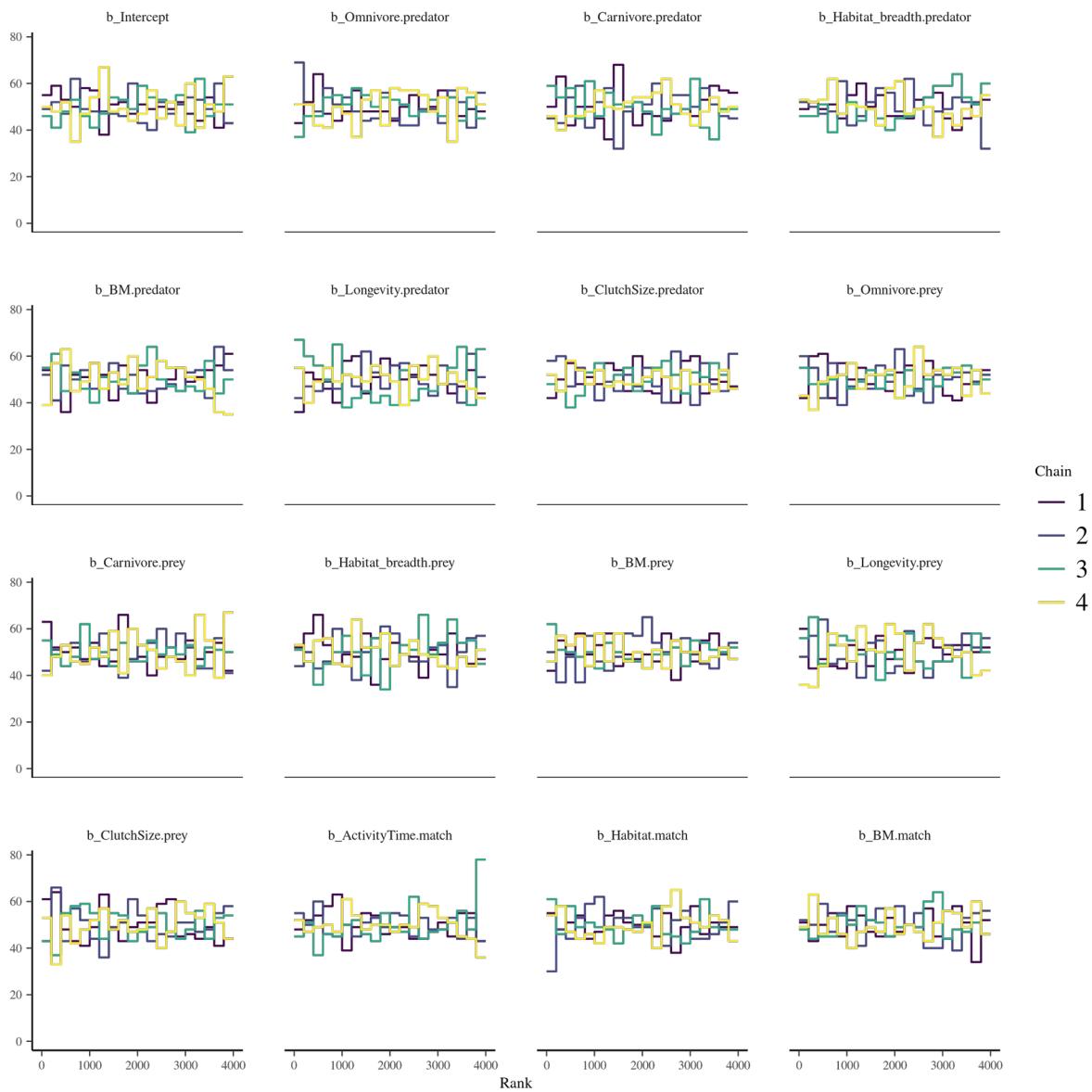
- Potential scale reduction factors:



*Figure S6.6: Distribution of the rank-normalized potential scale reduction factors on split chains for the food web of European terrestrial vertebrates.*

*Pyrenees model:*

- Rank plots of the population-level effects:



*Figure S6.7: Trace rank plots of the population-level effects for the model calibrated on the Pyrenees food web.*

- Posterior predictive checks:

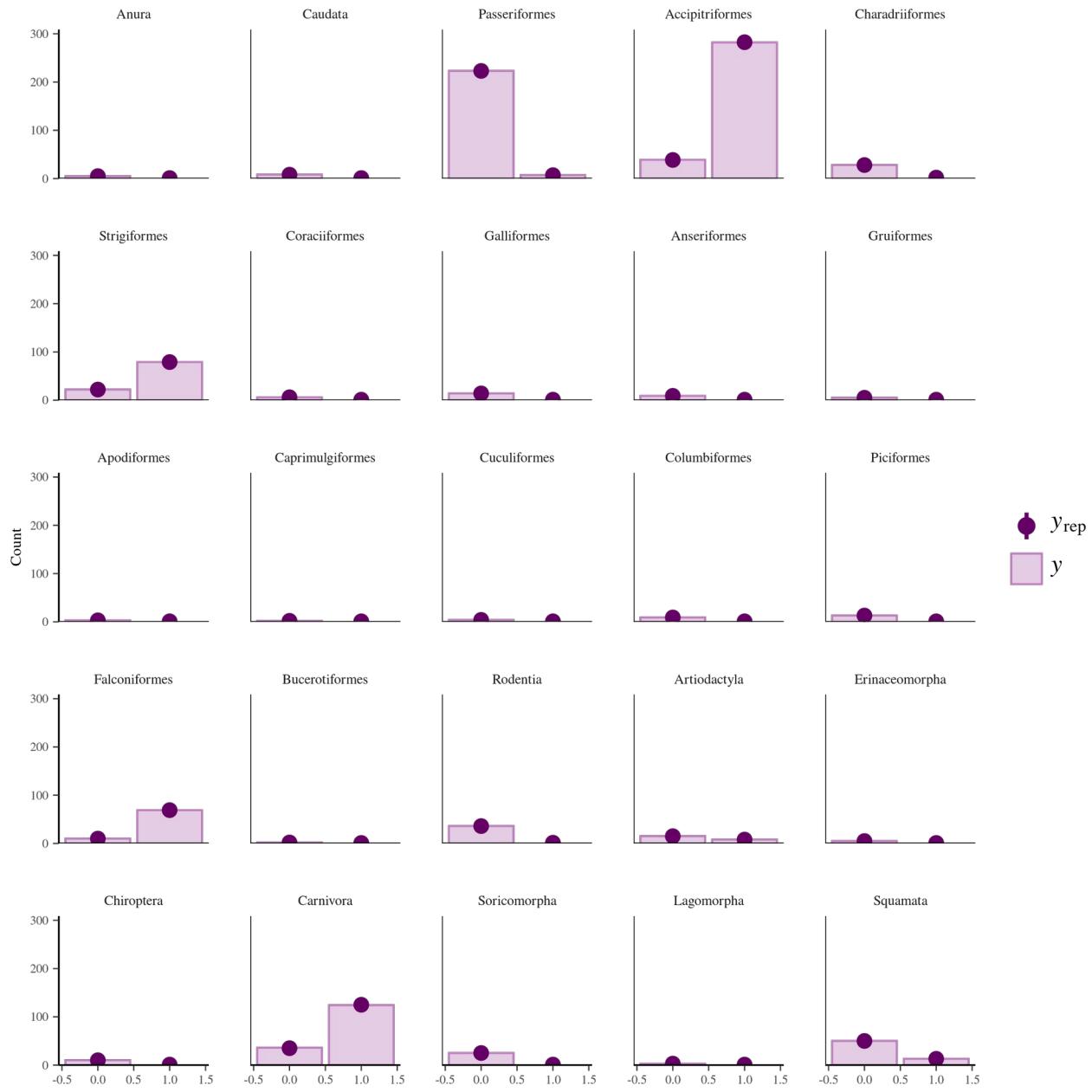
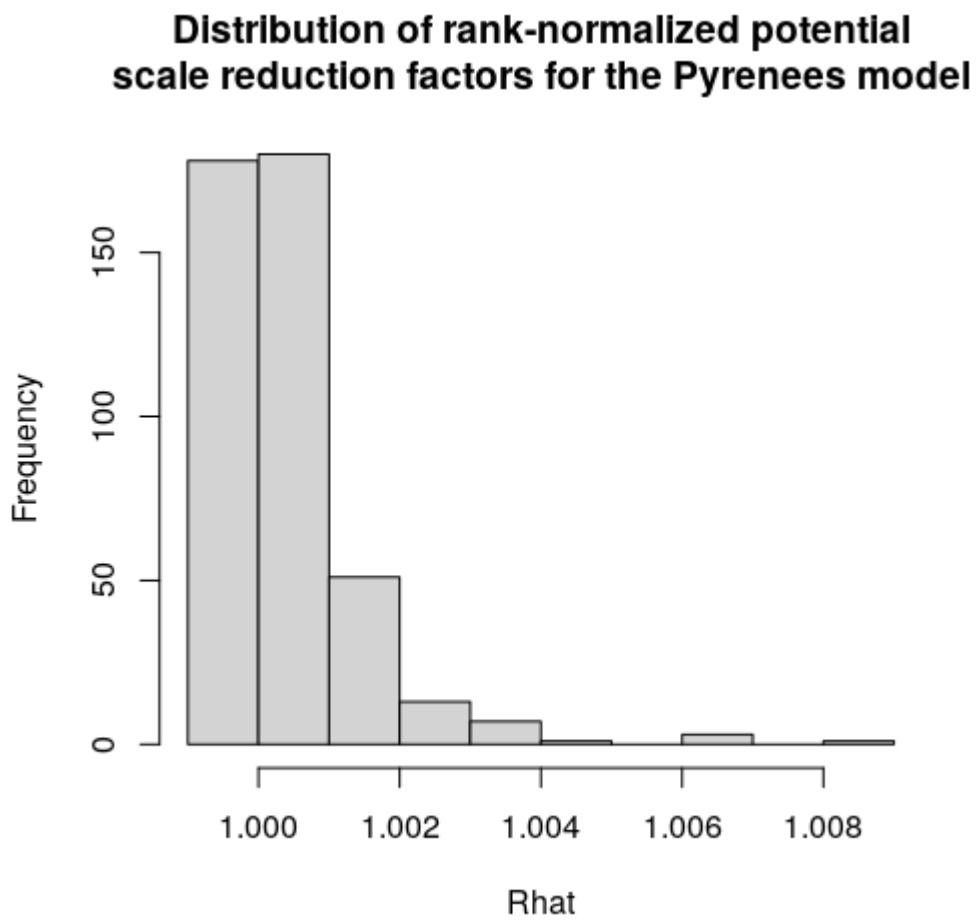


Figure S6.8: Comparison of the posterior predictive distribution ( $y_{rep}$ ; purple points) and the observed distribution ( $y$ ; purple bars) of the model calibrated on the Pyrenees food web.

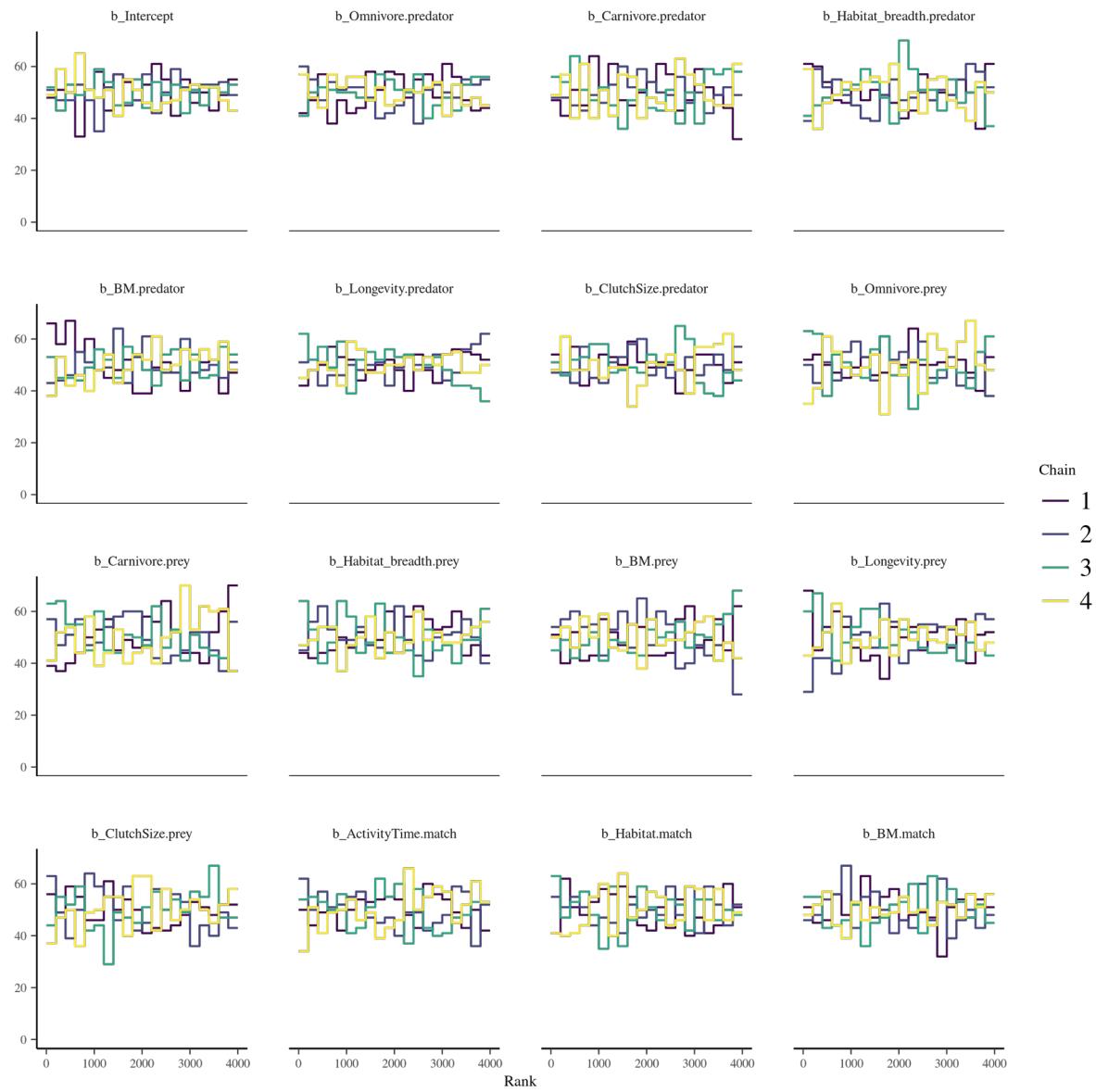
- Potential scale reduction factors:



*Figure S6.9: Distribution of the rank-normalized potential scale reduction factors on split chains for the Pyrenees food web*

### Serengeti model:

- Rank plots of the population-level effects:



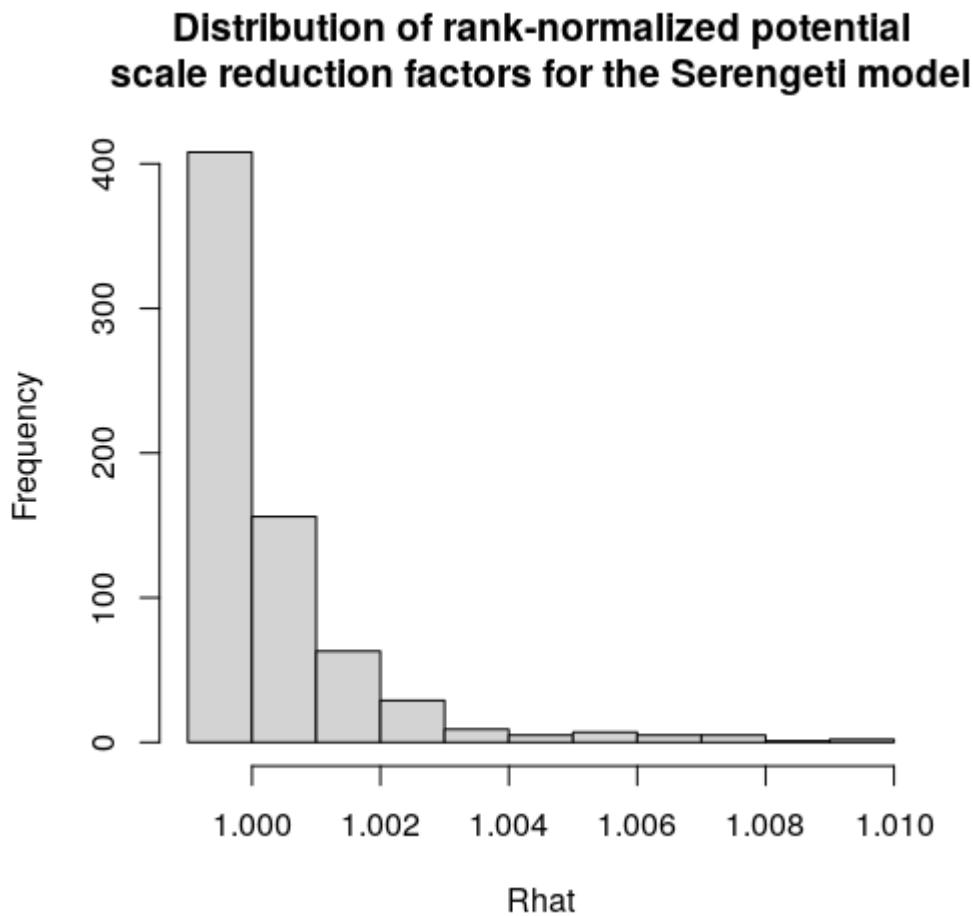
*Figure S6.10: Trace rank plots of the population-level effects for the model calibrated on the Serengeti food web.*

- Posterior predictive checks:



Figure S6.11: Comparison of the posterior predictive distribution ( $y_{rep}$ ; purple points) and the observed distribution ( $y$ ; purple bars) of the model calibrated on the Serengeti food web.

- Potential scale reduction factors:



*Figure S6.12: Distribution of the rank-normalized potential scale reduction factors on split chains for the Serengeti food web*

## Appendix S7: Predicting species interactions extended results

In this appendix, we present additional results on model performance. More precisely, we report additional performance metrics: the area under the precision-recall-gain curve, true positive rate, true negative rate, positive predictive value, and negative predictive value.

Like the area under the receiver operating characteristic curve (AUC), the **area under the precision-recall-gain curve (auprg)** is a threshold-independent metric that calculates the area under a curve. However, instead of using trade-off between false positive and true positive rate across thresholds, auprg uses the trade-off between positive predictive value (also called precision) and the true positive rate (also called recall) across thresholds. The auprg transform the precision-recall curve into an harmonic scale to fix some issues with the standard precision-recall curve (e.g., non-universality of baseline, non-linear interpolation). auprg varies between 0 for random predictions and 1 for perfect predictions. For more information, we suggest to refer to Flach & Kull (2015).

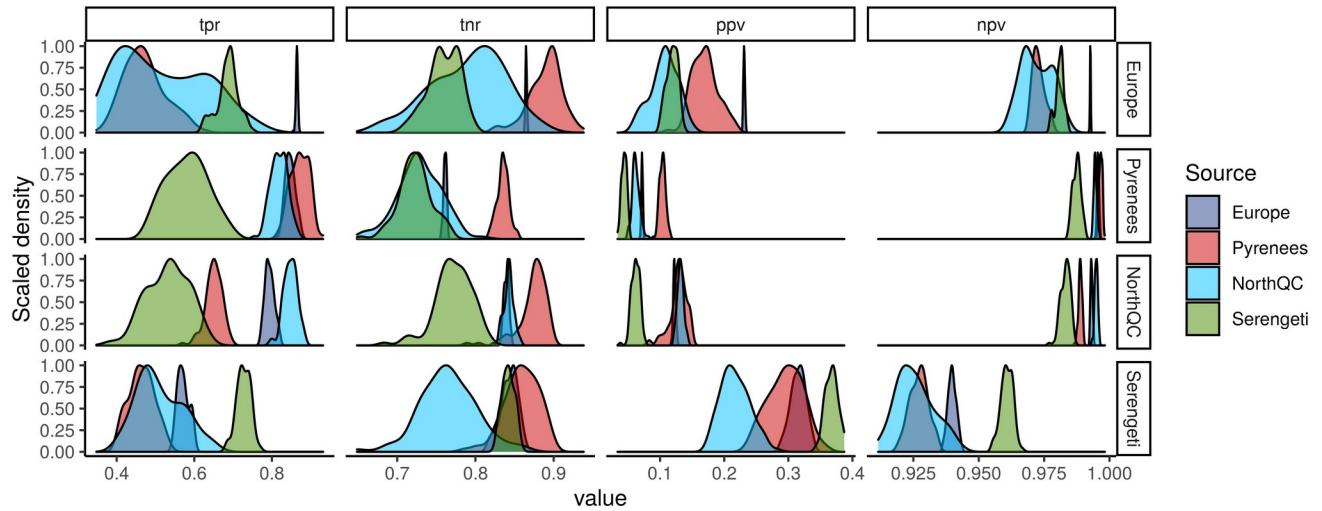
**True positive rate (tpr)** is the proportion of observed positives (interactions) predicted as positives (interactions). Instead of applying an arbitrary threshold on the predicted probabilities of interactions, we calculated the tpr for 100 samples of the posterior distribution.

**True negative rate (tnr)** is the proportion of observed negatives (non-interactions) predicted as negatives (non-interactions). Instead of applying an arbitrary threshold on the predicted probabilities of interactions, we calculated tnr for 100 samples of the posterior distribution.

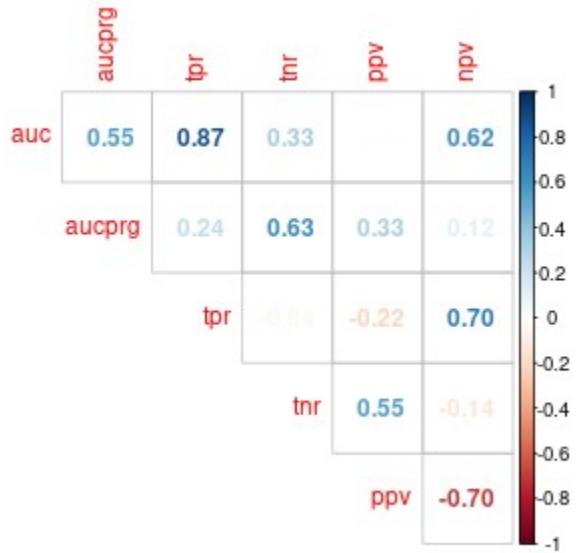
**Positive predictive value (ppv)** is the proportion of predicted positives (interactions) that are observed positives (interactions). It measures how confident we can be that a predicted interaction is an actual interaction. Instead of applying an arbitrary threshold on the predicted probabilities of interactions, we calculated ppv for 100 samples of the posterior distribution.

**Negative predictive value (npv)** is the proportion of predicted negatives (non-interactions) that are observed negatives (non-interactions). It measures how confident we can be that a predicted non-interaction is an actual non-interaction. Instead of applying an arbitrary threshold on the predicted

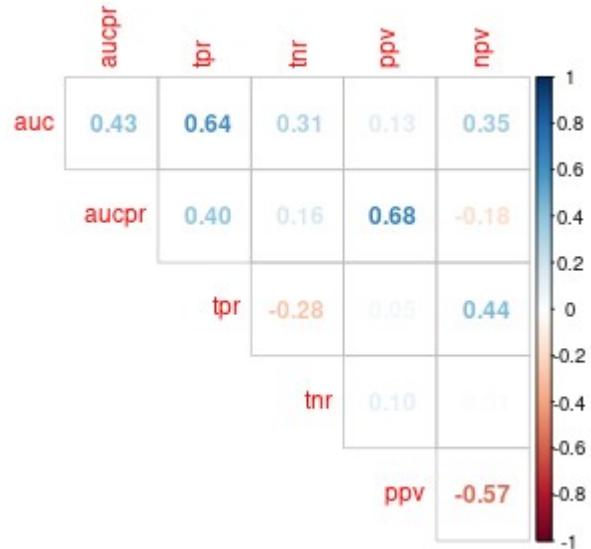
probabilities of interactions, we calculated the true negative rate for 100 samples of the posterior distribution.



*Figure S7.1: Scaled density distribution of true positive rate (tpr), true negative rate (tnr), positive predictive value (ppv), and negative predictive value (npv) of each model (Source) predicting every food webs (rows). tpr is the proportion of correctly predicted interactions, tnr is the proportion of correctly predicted non-interactions, ppv is the proportion of predicted interactions that are observed interactions, and npv is the proportion of predicted non-interactions that are observed non-interactions. We calculated tpr, tnr, ppv, and npv with 100 samples of the posterior distribution of the entire food webs.*



*Figure S7.2: Correlation between performance metrics for predicting the interactions of the entire food webs. auc is the area under the receiver operating curve, aucpr is the area under the precision-recall curve, tpr is the true positive rate, tnr is the true negative rate, ppv is the positive predictive value, and npv is the negative predictive value*



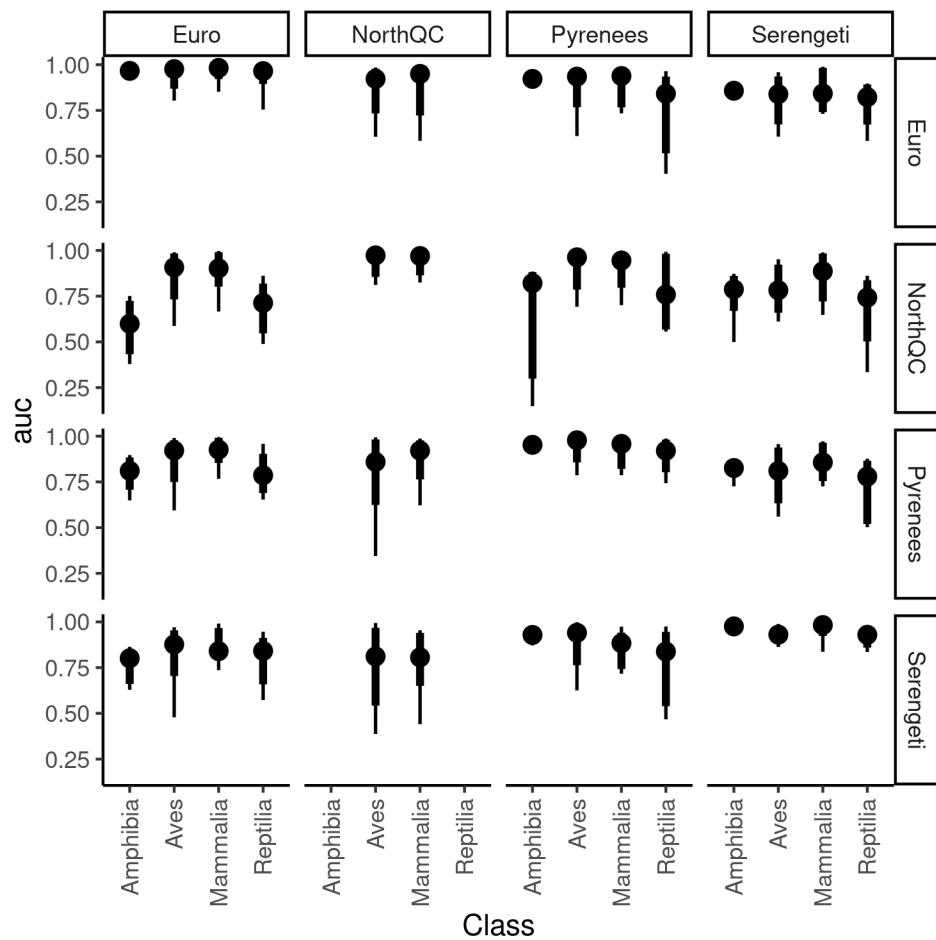
*Figure S7.3: Correlation between performance metrics for predicting the interactions of each species within a food web. See caption of figure S7.2 for abbreviations.*

## Model performance in relationship with distance:

Here, we present the outputs of the model relating model performance to geographic, phylogenetic, and environmental distances:

- Model 1: Total effect of geographic distance on AUC  
 $\text{logitauc} \sim \text{geo.dist\_sc} + (1|\text{Model}) + (1|\text{Food web})$   
fixed effects:  
Intercept: Estimate = 1.83, 95% CrI = [1.24, 2.26]  
 $\text{geo.dist\_sc}$ : Estimate = -0.57, 95% CrI = [-0.78, -0.36]  
random effect:  
 $\text{sd}(1|\text{Model})$ : Estimate = 0.25, 95% CrI = [0.01, 0.91]  
 $\text{sd}(1|\text{Food web})$ : Estimate = 0.27, 95% CrI = [0.01, 1.11]
- Model 2: Partial effect of geographic distance on AUC  
 $\text{logitauc} \sim \text{geo.dist\_sc} + \text{phylo.dist\_sc} + \text{env.dist\_sc} + (1|\text{Model}) + (1|\text{Food web})$   
fixed effects:  
Intercept: Estimate = 1.78, 95% CrI = [0.70, 2.25]  
 $\text{geo.dist\_sc}$ : Estimate = -0.26, 95% CrI = [-0.87, 0.35]  
 $\text{phylo.dist\_sc}$ : Estimate = -0.29, 95% CrI = [-0.74, 0.13]  
 $\text{env.dist\_sc}$ : Estimate = -0.10, 95% CrI = [-0.74, 0.60]  
random effect:  
 $\text{sd}(1|\text{Model})$ : Estimate = 0.39, 95% CrI = [0.01, 2.06]  
 $\text{sd}(1|\text{Food web})$ : Estimate = 0.24, 95% CrI = [0.01, 1.01]
- Model 3: Effect of phylogenetic distance on AUC after controlling for geographic distance  
 $\text{logitauc} \sim \text{phylo.dist\_sc} + \text{geo.dist\_sc} + (1|\text{Source}) + (1|\text{Target})$   
fixed effects:  
Intercept: Estimate = 1.82, 95% CrI = [1.13, 2.24]  
 $\text{geo.dist\_sc}$ : Estimate = -0.34, 95% CrI = [-0.68, 0.01]  
 $\text{phylo.dist\_sc}$ : Estimate = -0.31, 95% CrI = [-0.69, 0.06]  
random effect:  
 $\text{sd}(1|\text{Model})$ : Estimate = 0.30, 95% CrI = [0.01, 1.07]  
 $\text{sd}(1|\text{Food web})$ : Estimate = 0.24, 95% CrI = [0.01, 0.98]
- Model 4: Effect of environmental distance on AUC after controlling for geographic distance  
 $\text{logitauc} \sim \text{env.dist\_sc} + \text{geo.dist\_sc} + (1|\text{Source}) + (1|\text{Target})$   
fixed effects:  
Intercept: Estimate = 1.83, 95% CrI = [1.12, 2.24]  
 $\text{geo.dist\_sc}$ : Estimate = -0.29, 95% CrI = [-0.90, 0.31]  
 $\text{env.dist\_sc}$ : Estimate = -0.30, 95% CrI = [-0.90, 0.32]  
random effect:  
 $\text{sd}(1|\text{Model})$ : Estimate = 0.23, 95% CrI = [0.01, 0.92]  
 $\text{sd}(1|\text{Food web})$ : Estimate = 0.25, 95% CrI = [0.01, 1.12]

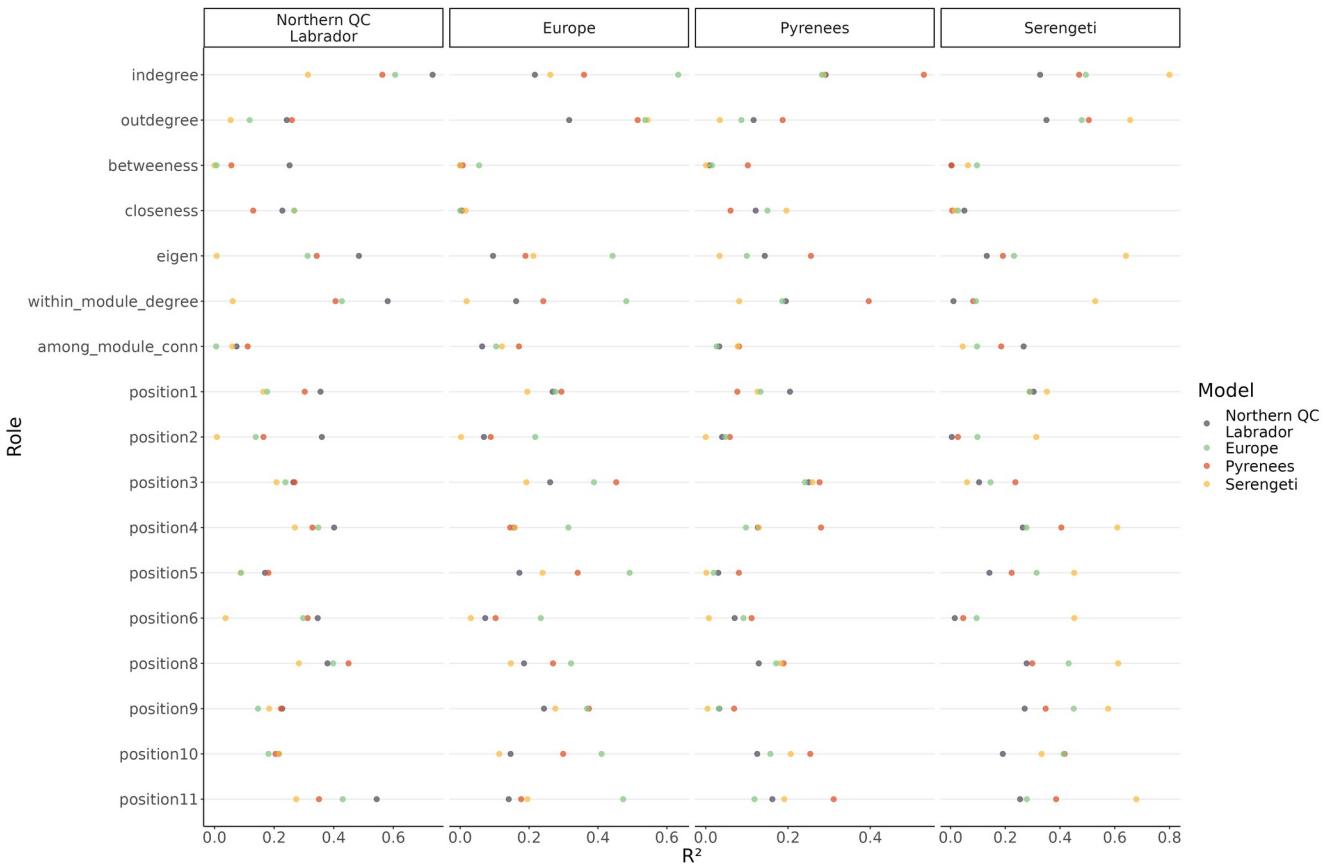
## Model performance in relationship with species class:



*Figure S7.4: Class specific performance of predicting species interaction for each combination of model (rows) and food web predicted (columns). The point intervals represent the median, the 80%, and 95% intervals of auc across species within each class.*

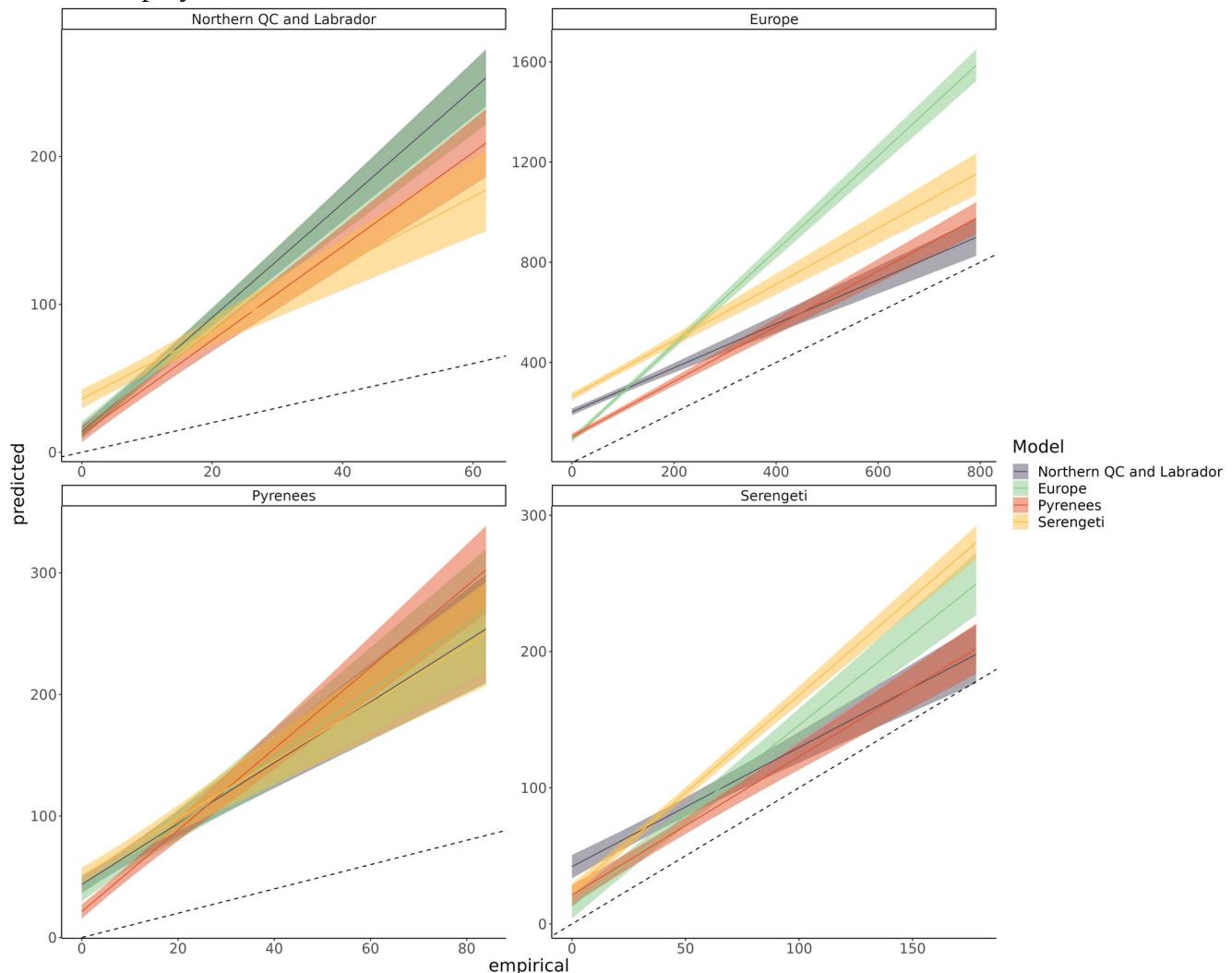
## **Appendix S8: Predicting species role extended results**

For each of the combination of species role metric considered, model and food web predicted, we fitted a simple linear regression using OLS through the function `lm()` in R. The response variable is the predicted values, and the predictor variables are the observed values. We extracted the slopes, intercept, and  $R^2$  of each regression.  $R^2$  measures the correlation between the predicted and observed values. The intercept measure the average biases:  $\text{intercept} > 0$  represent average overestimation,  $\text{intercept} < 0$  means average underestimation. The slope measures the homogenization or heterogenization of the values:  $\text{slope} > 1$  means species have more variable values in the predictions,  $\text{slope} < 1$  means that species have less variable values in the predictions. The following figures show the  $R^2$  (Figure S8.1) and the fitted line of the linear regression (Figure S8.2-20) to ease the interpretation of the results.



*Figure S8.1: Predictive performance of each model (colors) to predict species role for every food webs (column). Performance is measured with the coefficient of determination ( $R^2$ ). From top to bottom: Number of prey, number of predator, betweenness, closeness, eigenvector centrality, within-module degree, participation coefficient, frequency of predator position in a linear chain motif, consumer position in a linear chain motif, resource in a linear chain motif, omnivore in a intraguild predation motif, consumer in a intraguild predation motif, resource in a intraguild predation motif, predation in a direct competition motif, prey in a direct competition motif, predator in a apparent competition motif, and prey in a apparent competition motif.*

## Number of prey



*Figure S8.2: Linear regression comparing the predicted number of prey and empirical number of prey of species for each model predicting every food webs. The main bias is that the number of prey is overpredicted (intercept > 0). In a many cases, the bias is more important for generalist predators (slope > 1). Interestingly, the Serengeti model tend to better predict generalist species, but less specialist species (compared to other models). It's overall performance is still much lower for predicting the number of prey in the Northern Quebec food web (Figure S8.1) given there are many more specialists than generalists.*

## Number of predators

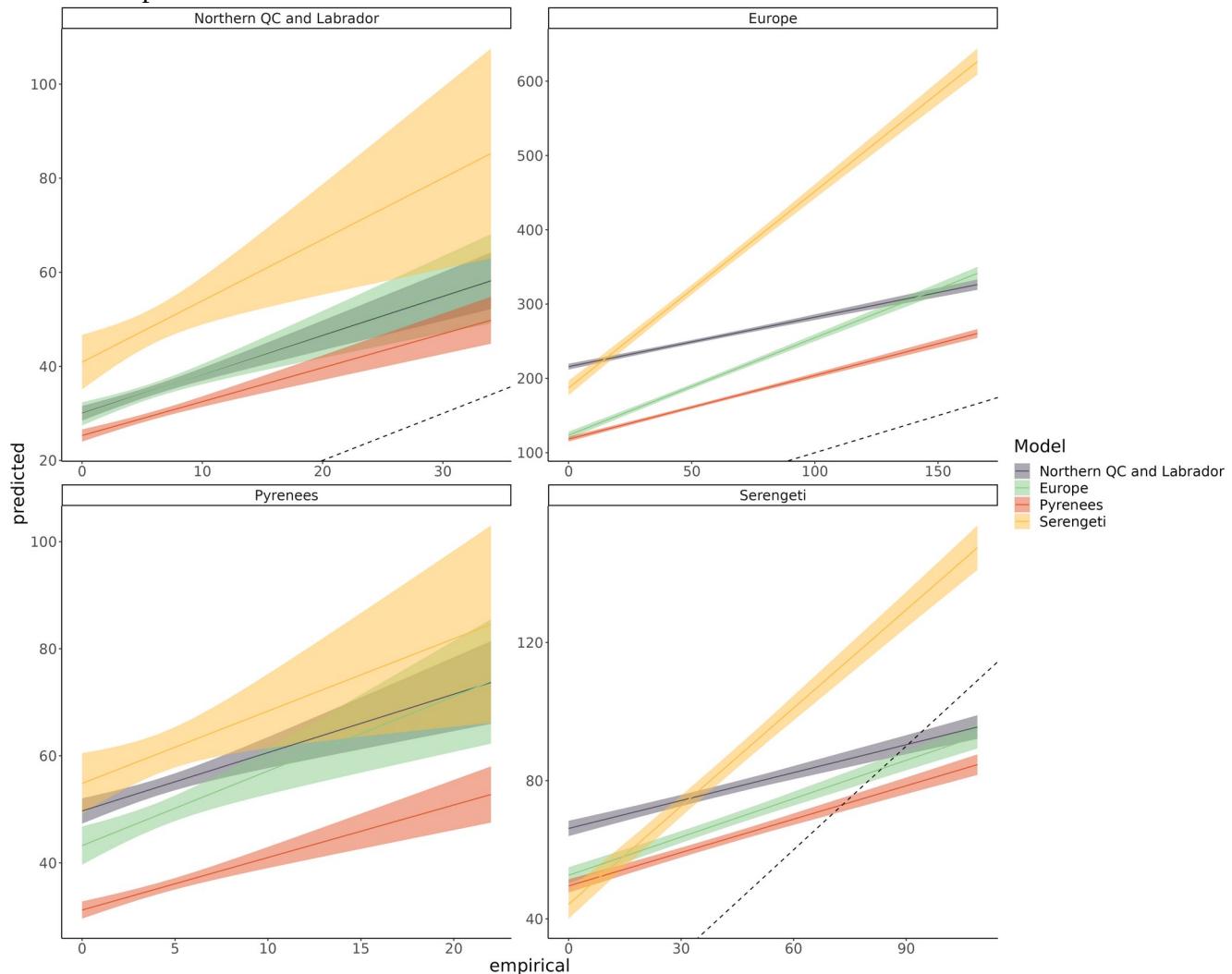
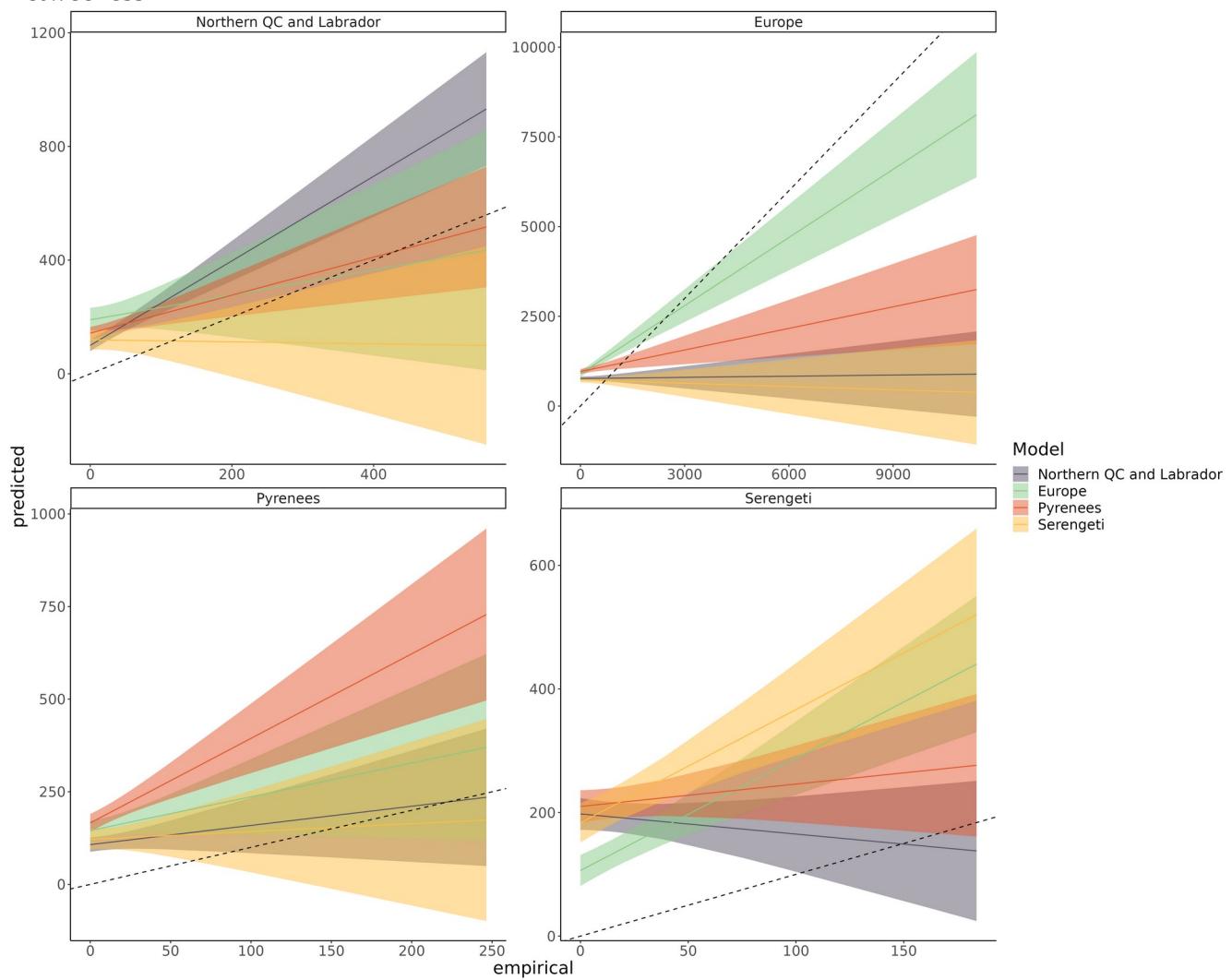


Figure S8.3: Linear regression comparing the predicted number of predators and empirical number of predators of species for each model predicting every food webs. The main bias is that the number of prey is overpredicted (intercept > 0). In many cases, the bias is more important for generalist predators ( $slope < 1$ ).

## Betweenness



*Figure S8.4: Linear regression comparing the predicted betweenness centrality and empirical betweenness centrality of species for each model predicting every food webs. The main bias is that predicted betweenness is more homogeneous across species (slope < 1; overpredicted for low end of the range, and underpredicted for high end of the range).*

## Closeness

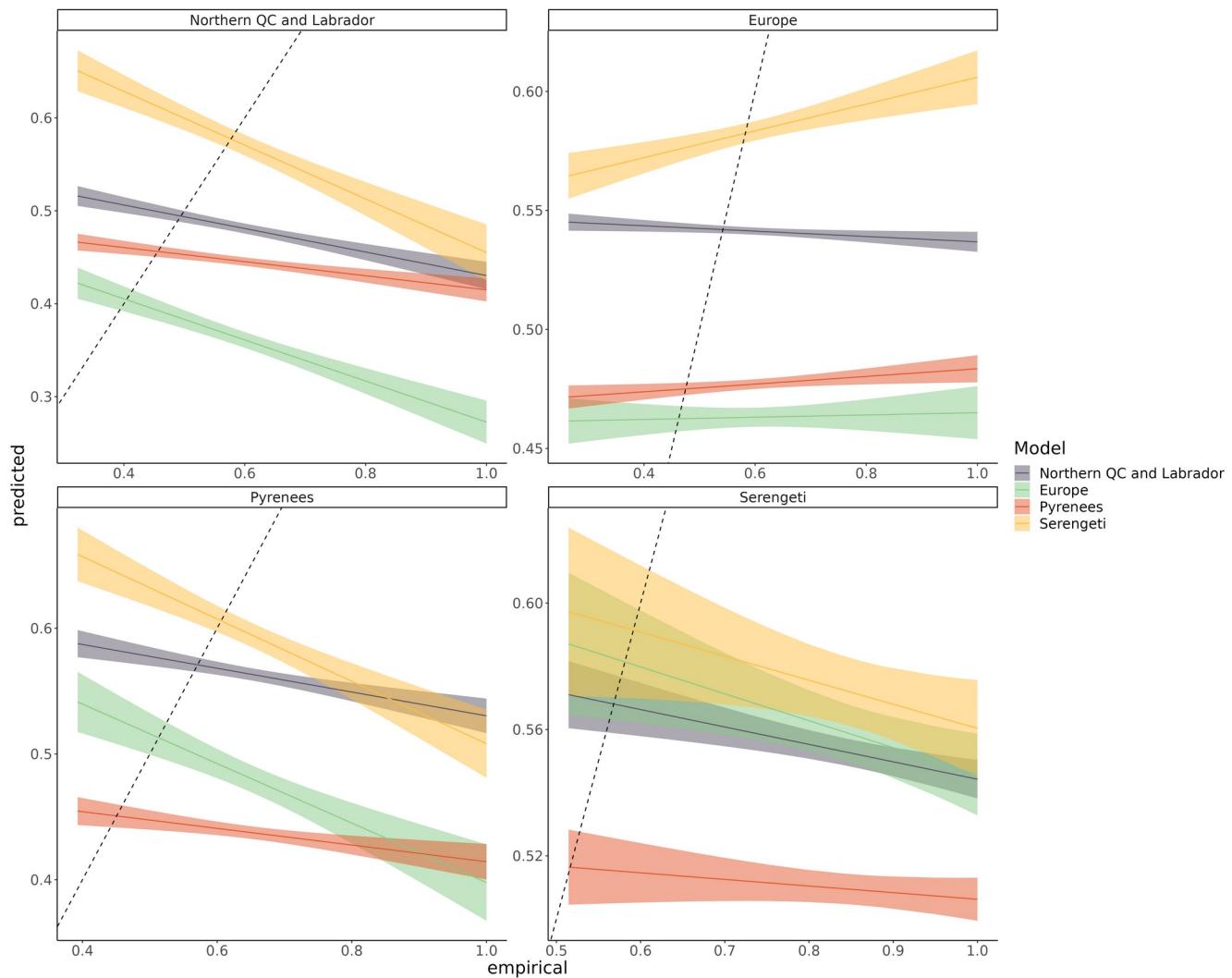
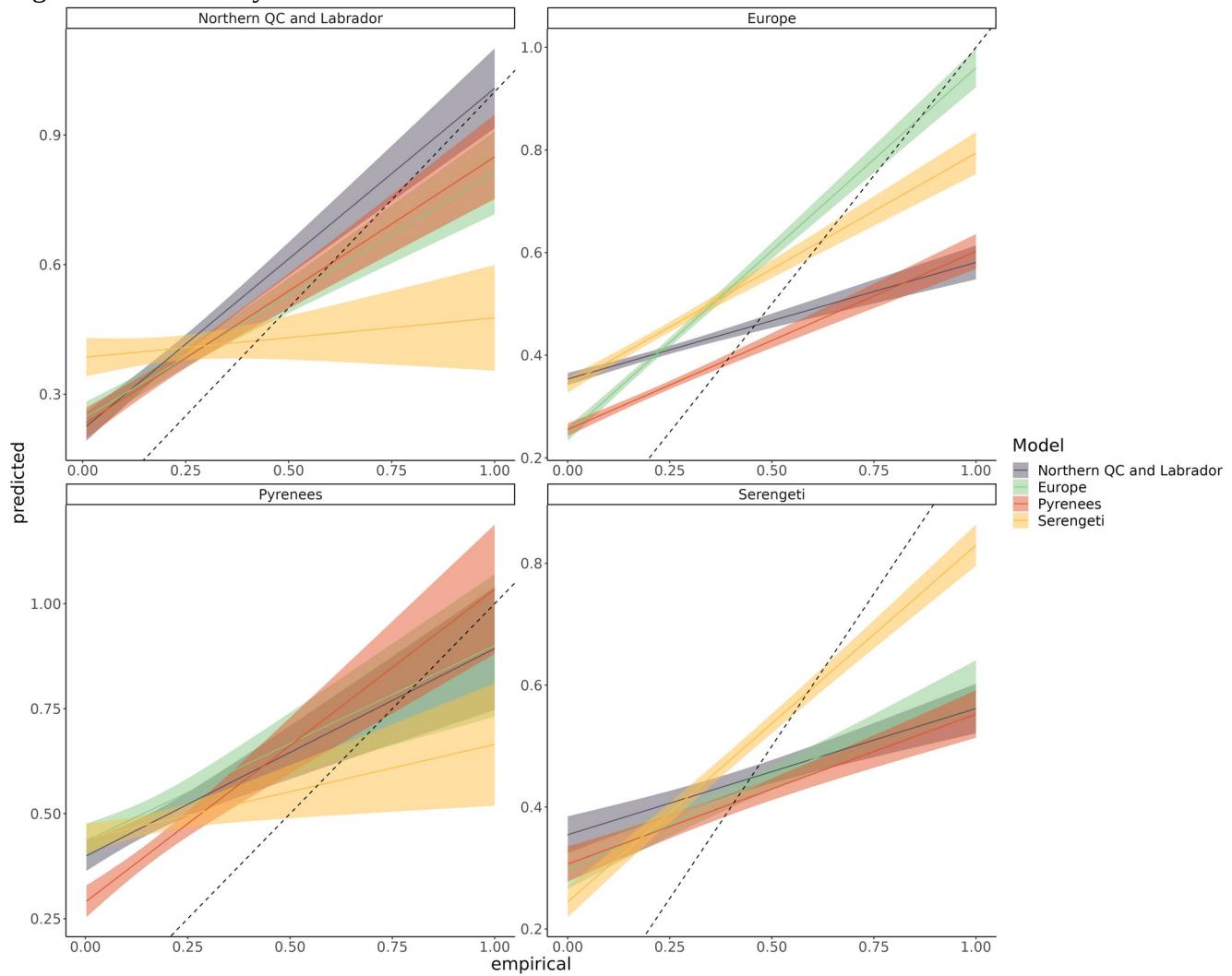


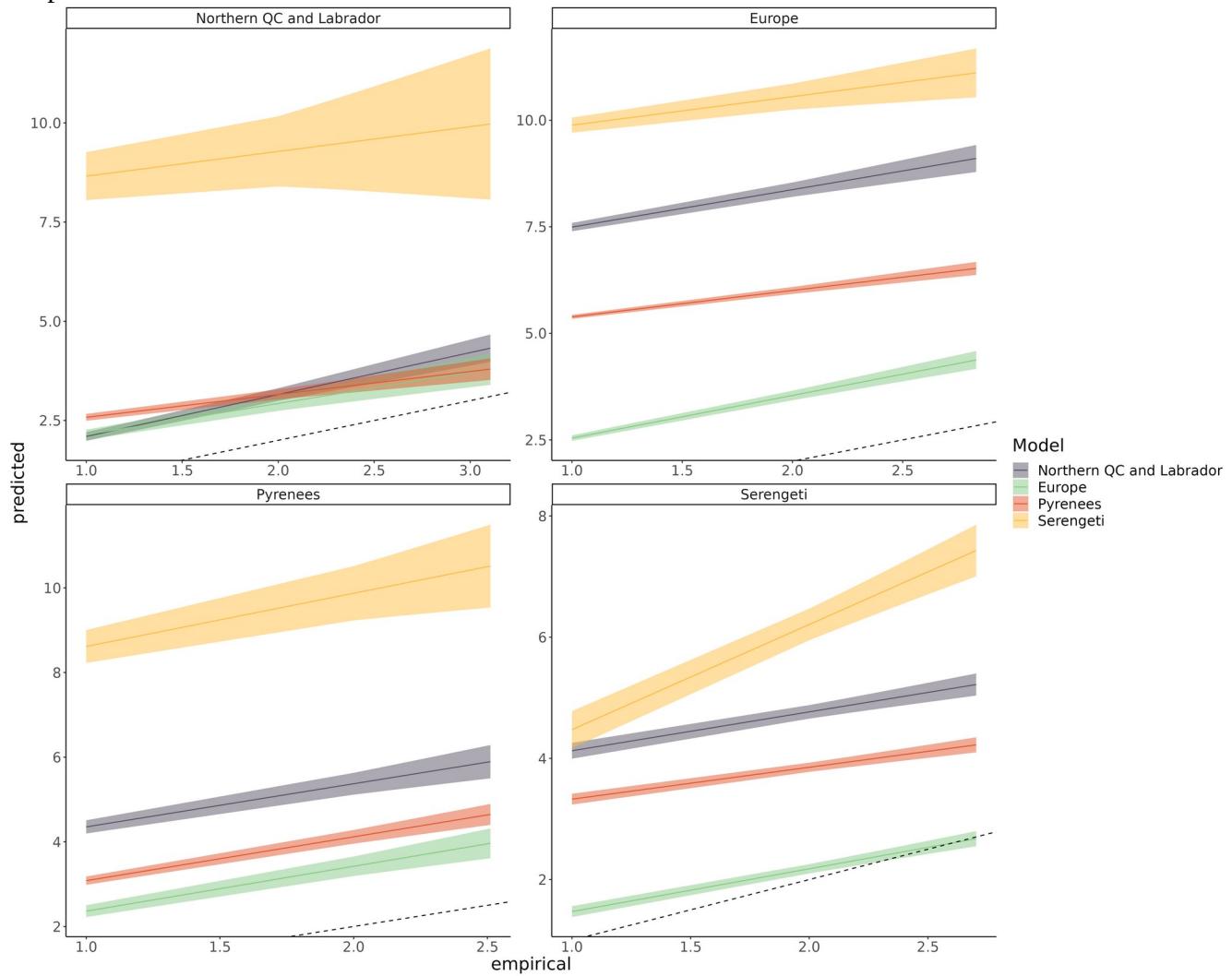
Figure S8.5: Linear regression comparing the predicted closeness centrality and empirical closeness centrality of species for each model predicting every food webs. The main bias is that predicted closeness is more homogeneous across species (slope < 1; overpredicted for low end of the range, and underpredicted for high end of the range).

## Eigenvector centrality



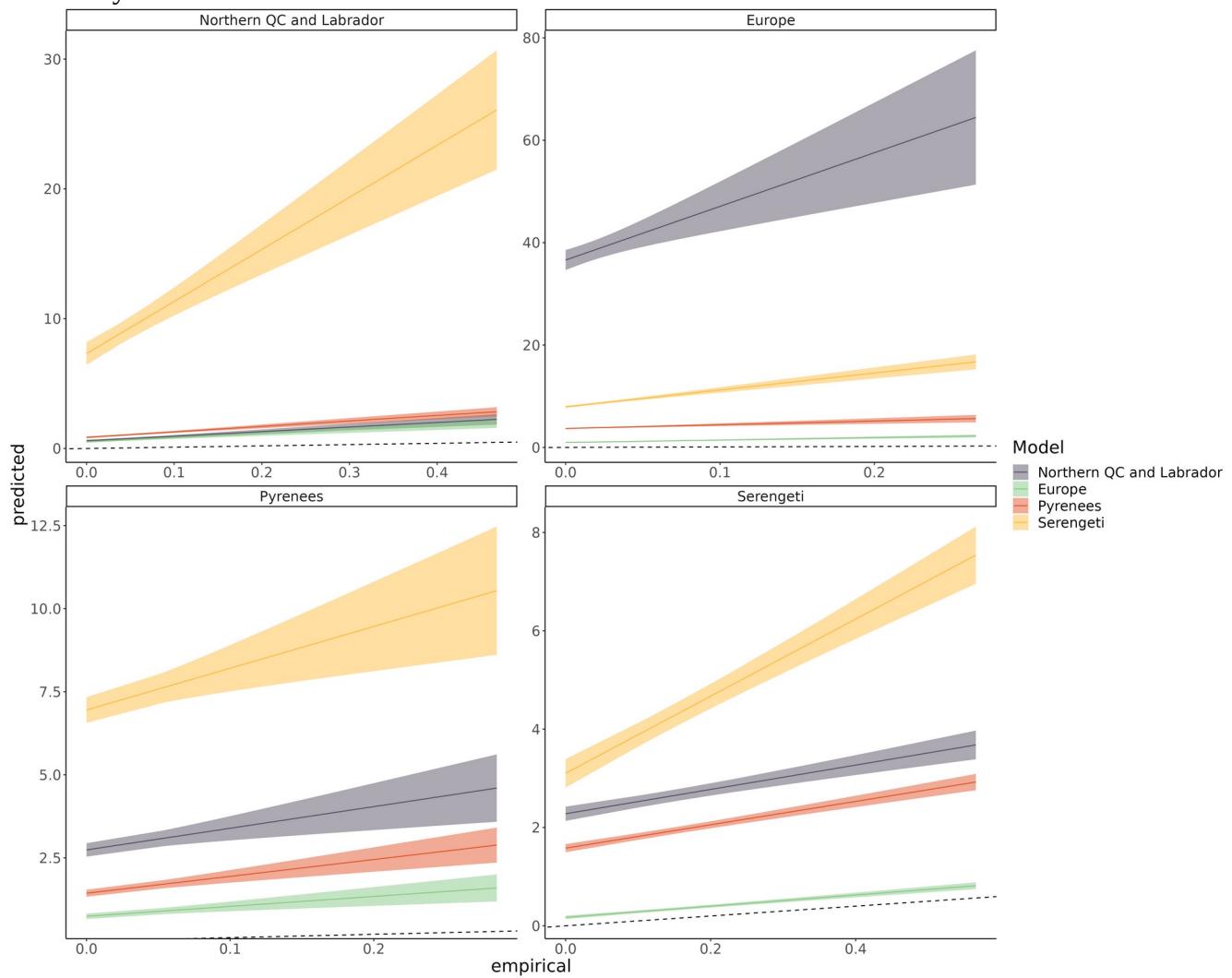
*Figure S8.6: Linear regression comparing the predicted eigenvector centrality and empirical eigenvector centrality of species for each model predicting every food webs. The main bias is that predicted eigenvector centrality is more homogeneous across species (slope < 1; overpredicted for low end of the range, and underpredicted for high end of the range).*

## Trophic Level



*Figure S8.7: Linear regression comparing the predicted trophic level and empirical trophic level of species for each model predicting every food webs. The main bias is that predicted the trophic level of most species is overpredicted across the range (intercept > 0).*

## Omnivory



*Figure S8.8: Linear regression comparing the predicted omnivory index and empirical omnivory index of species for each model predicting every food webs. The main bias is that predicted the level of omnivory of most species is overpredicted across the range (intercept > 0).*

### Within-module degree

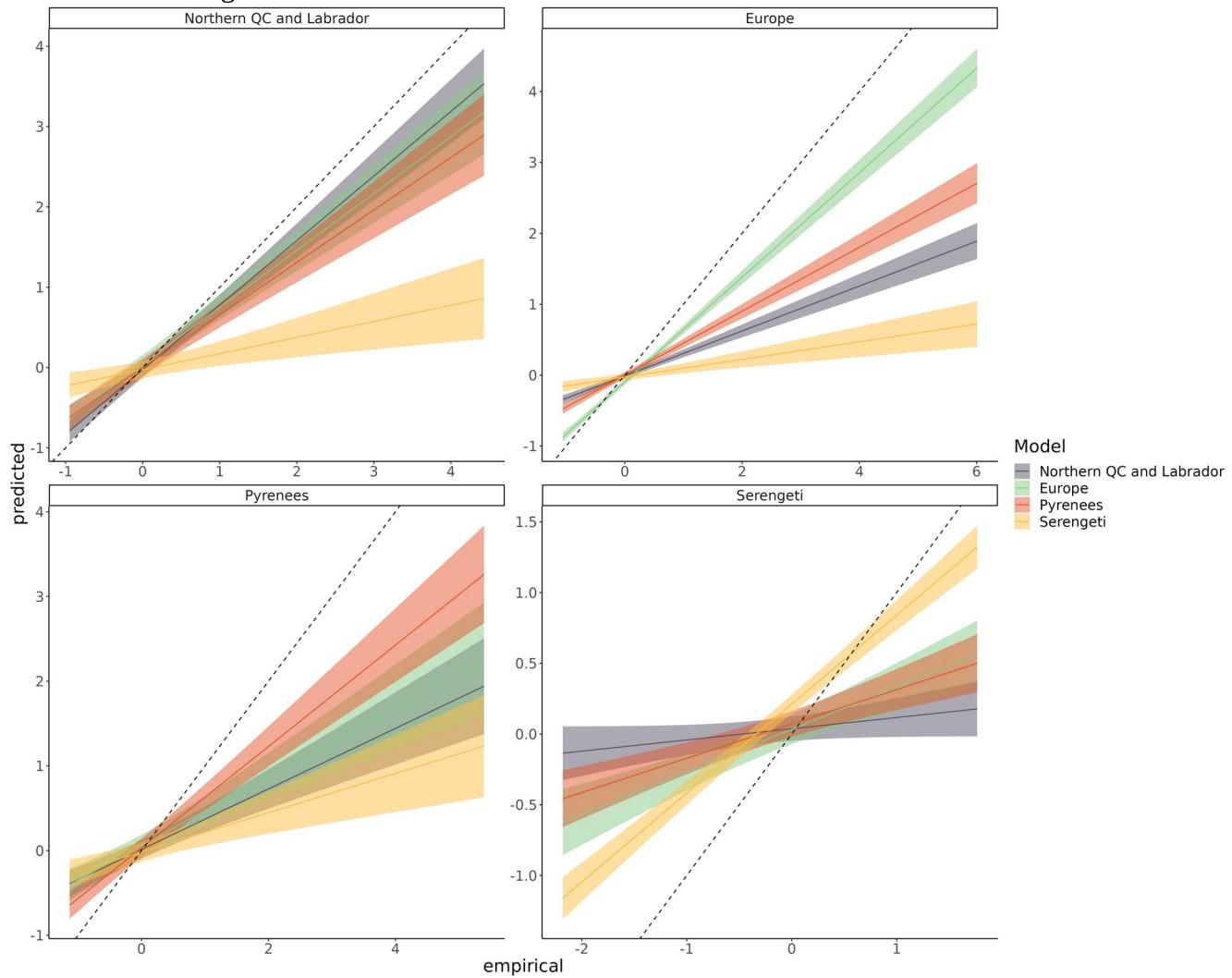


Figure S8.9: Linear regression comparing the predicted within-module degree and empirical within-module degree of species for each model predicting every food webs. The main bias is that predicted within-module degree is more homogeneous across species (slope < 1; overpredicted for low end of the range, and underpredicted for high end of the range).

## Participation coefficient

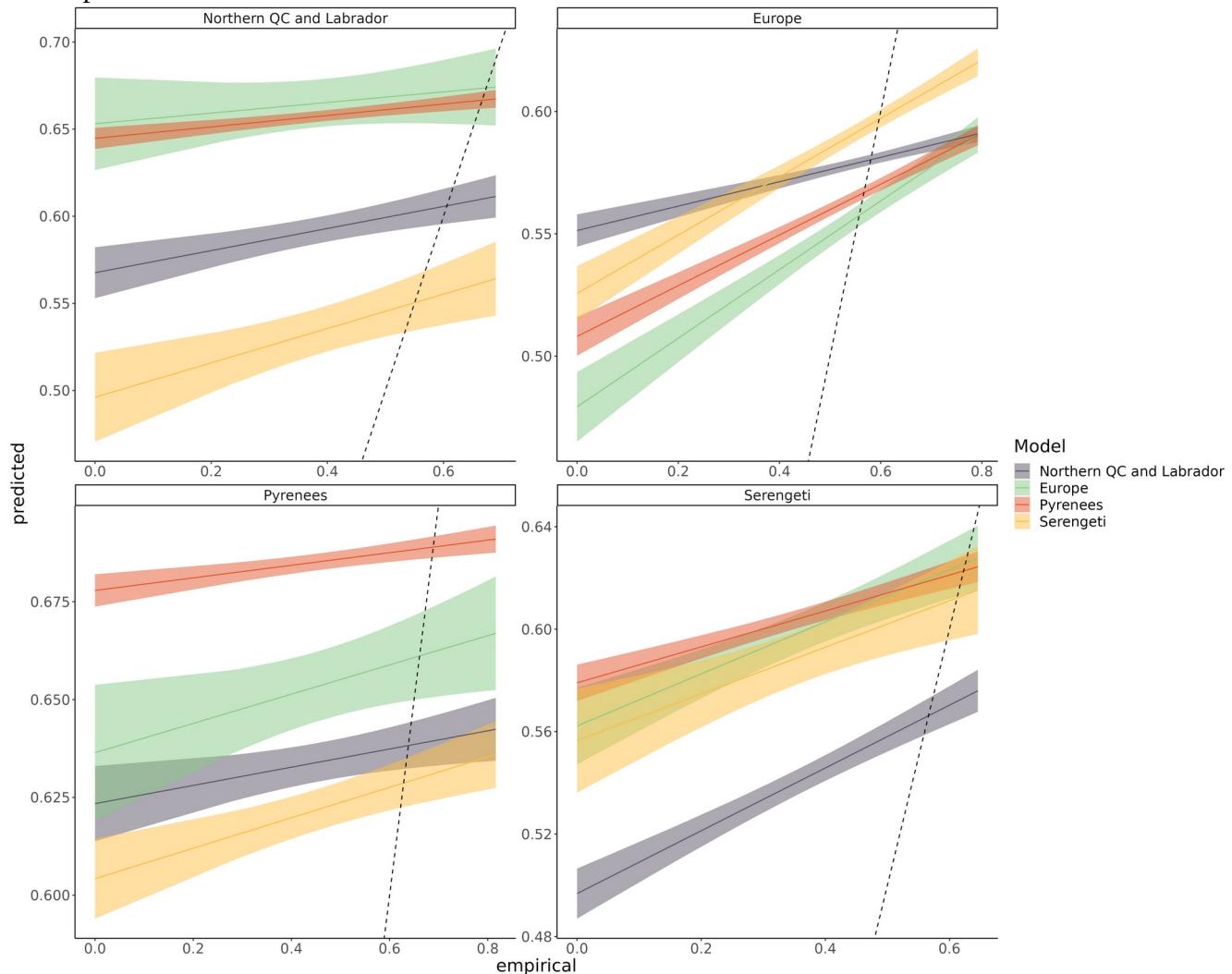
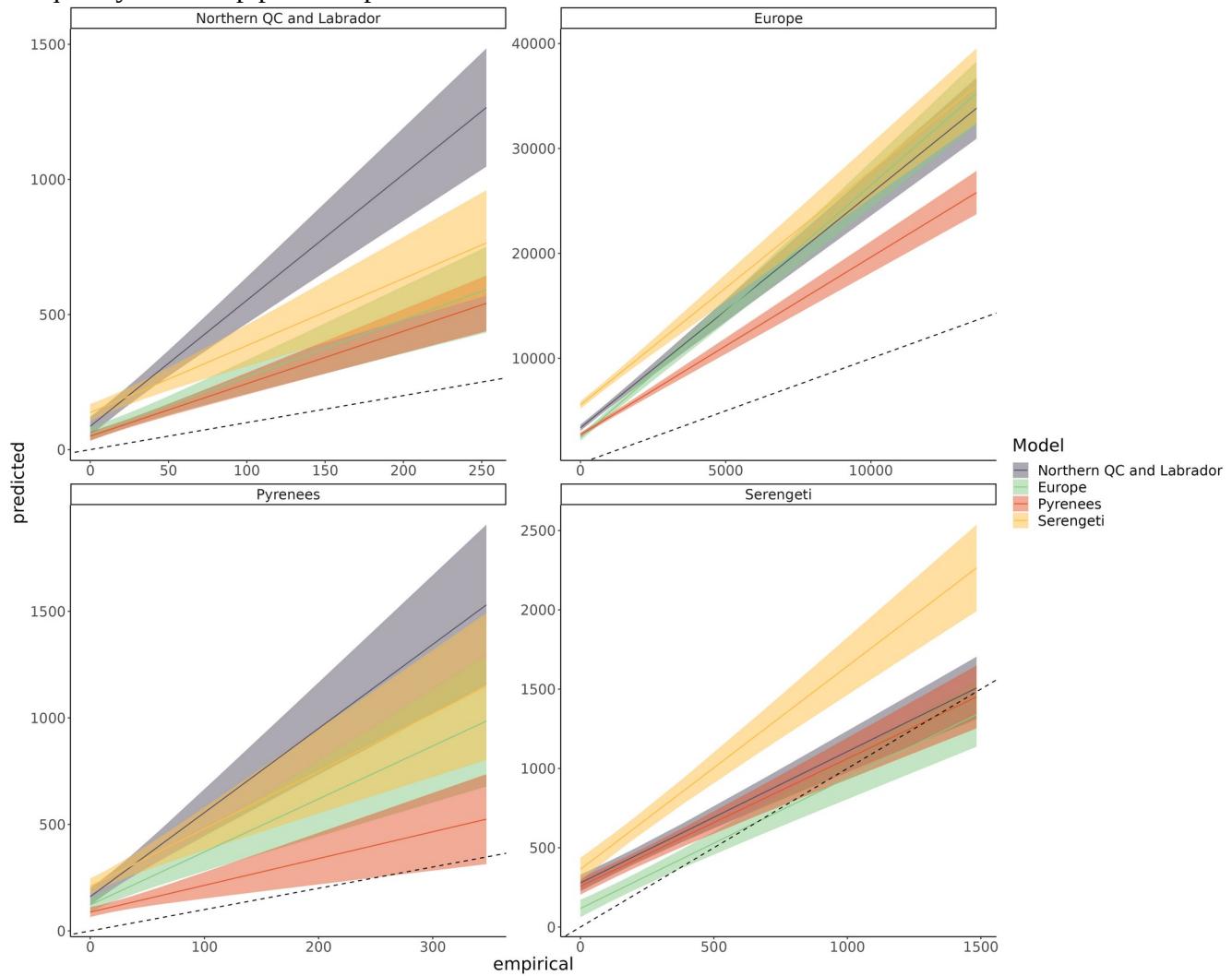


Figure S8.10: Linear regression comparing the predicted participation coefficient and empirical participation coefficient of species for each model predicting every food webs. The main bias is that predicted participation coefficient is more homogeneous across species (slope < 1; overpredicted for low end of the range, and underpredicted for high end of the range).

### Frequency of the top predator position in the linear food chain motif



*Figure S8.11: Linear regression comparing the predicted frequency of the top predator position in the linear food chain motif and empirical frequency of the top predator position in the linear food chain motif of species for each model predicting every food webs. The main bias is that predicted the frequency of top predator position of most species is overpredicted across the range (intercept > 0).*

### Frequency of the consumer position in the linear food chain motif

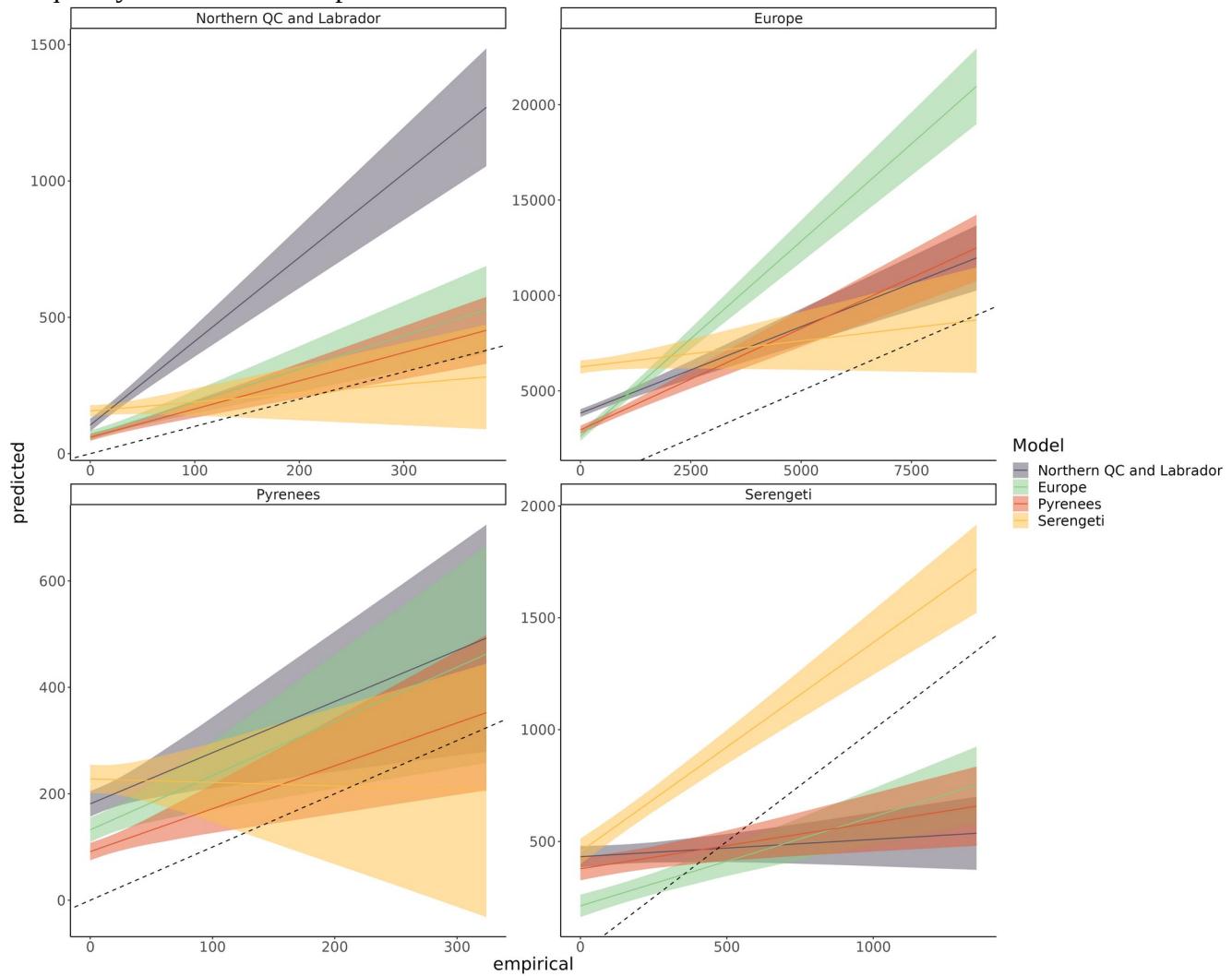
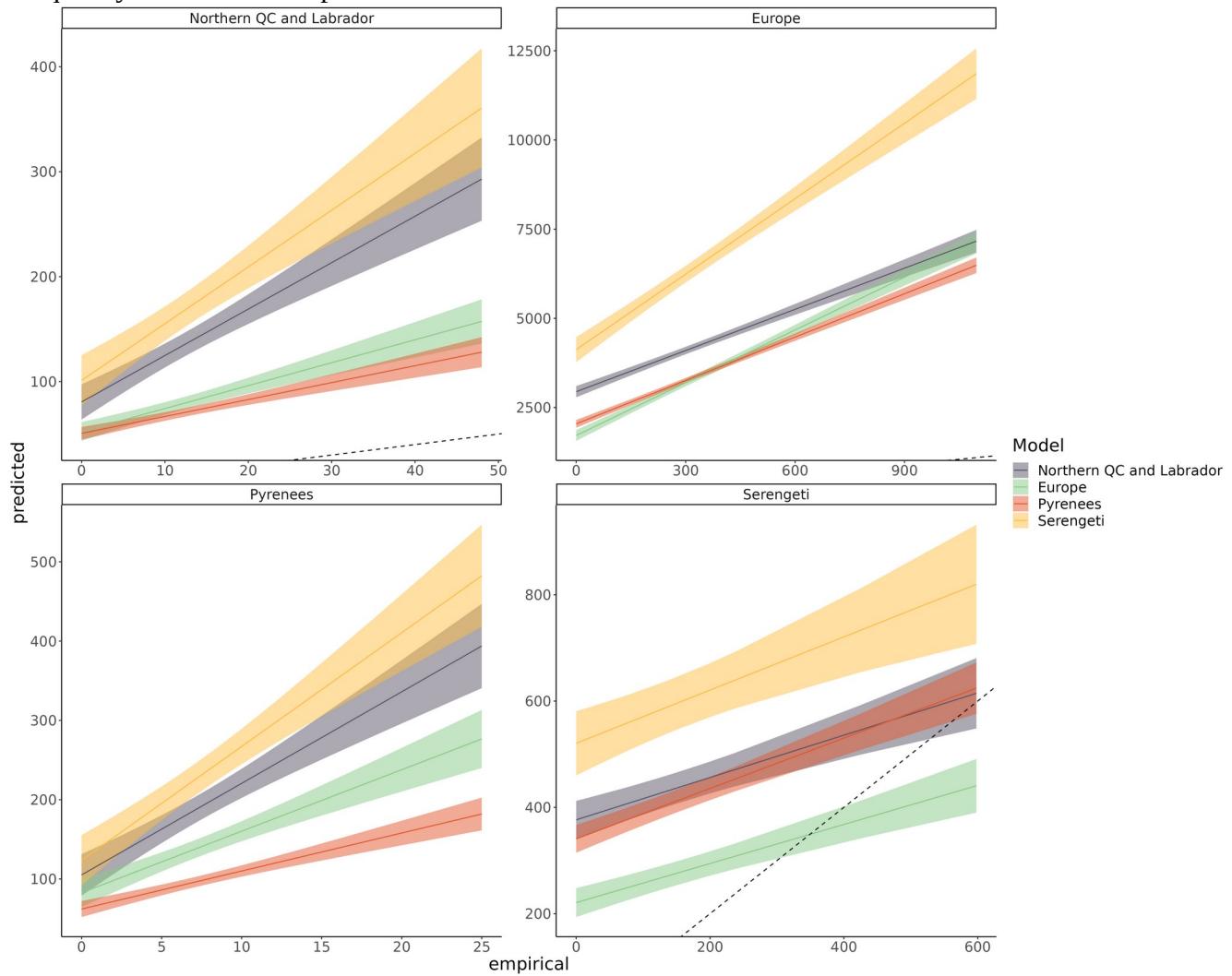


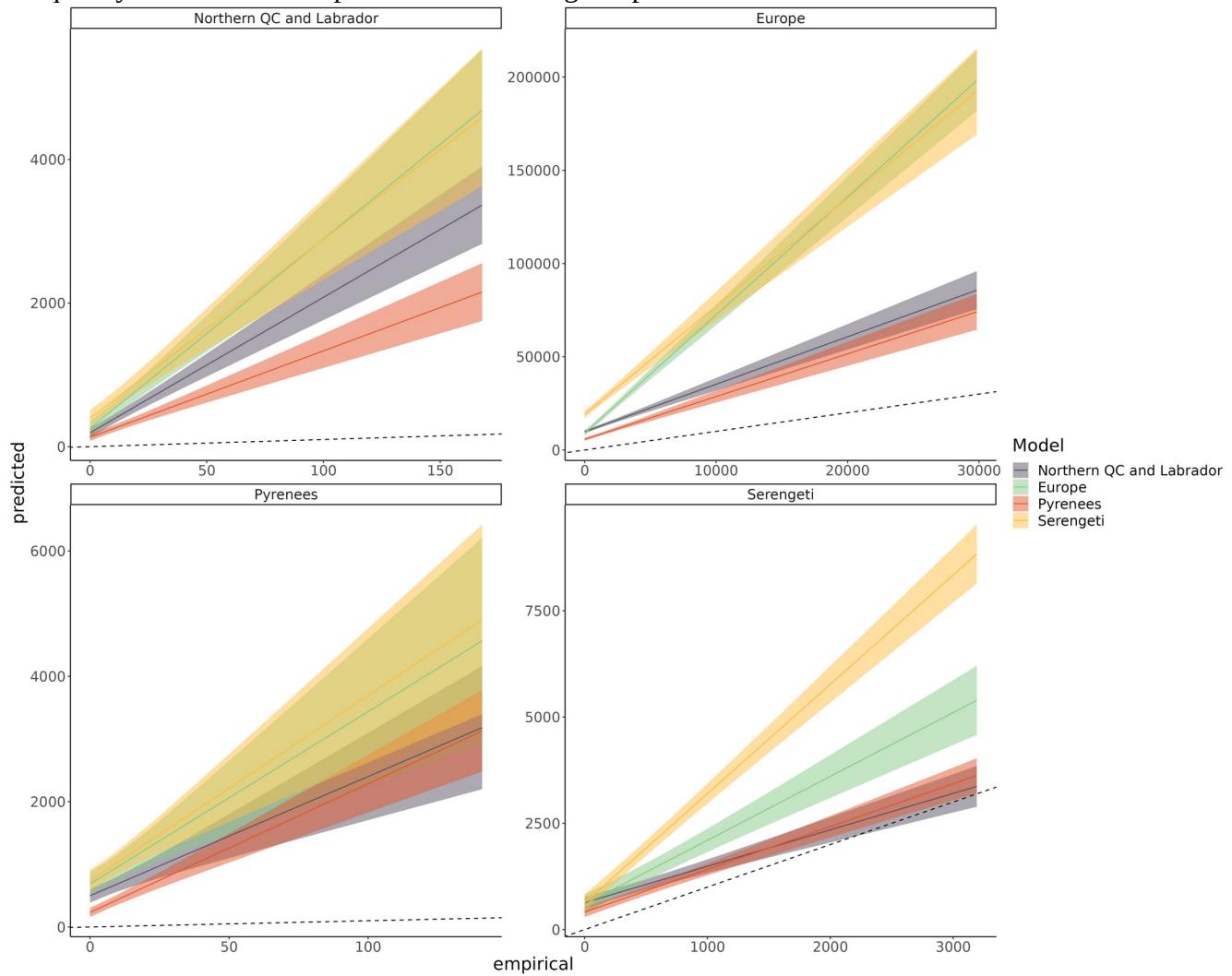
Figure S8.12: Linear regression comparing the predicted frequency of the consumer position in the linear food chain motif and empirical frequency of the consumer position in the linear food chain motif of species for each model predicting every food webs. The main bias is that predicted the frequency of consumer position of most species is overpredicted across the range (intercept > 0).

### Frequency of the resource position in the linear food chain motif



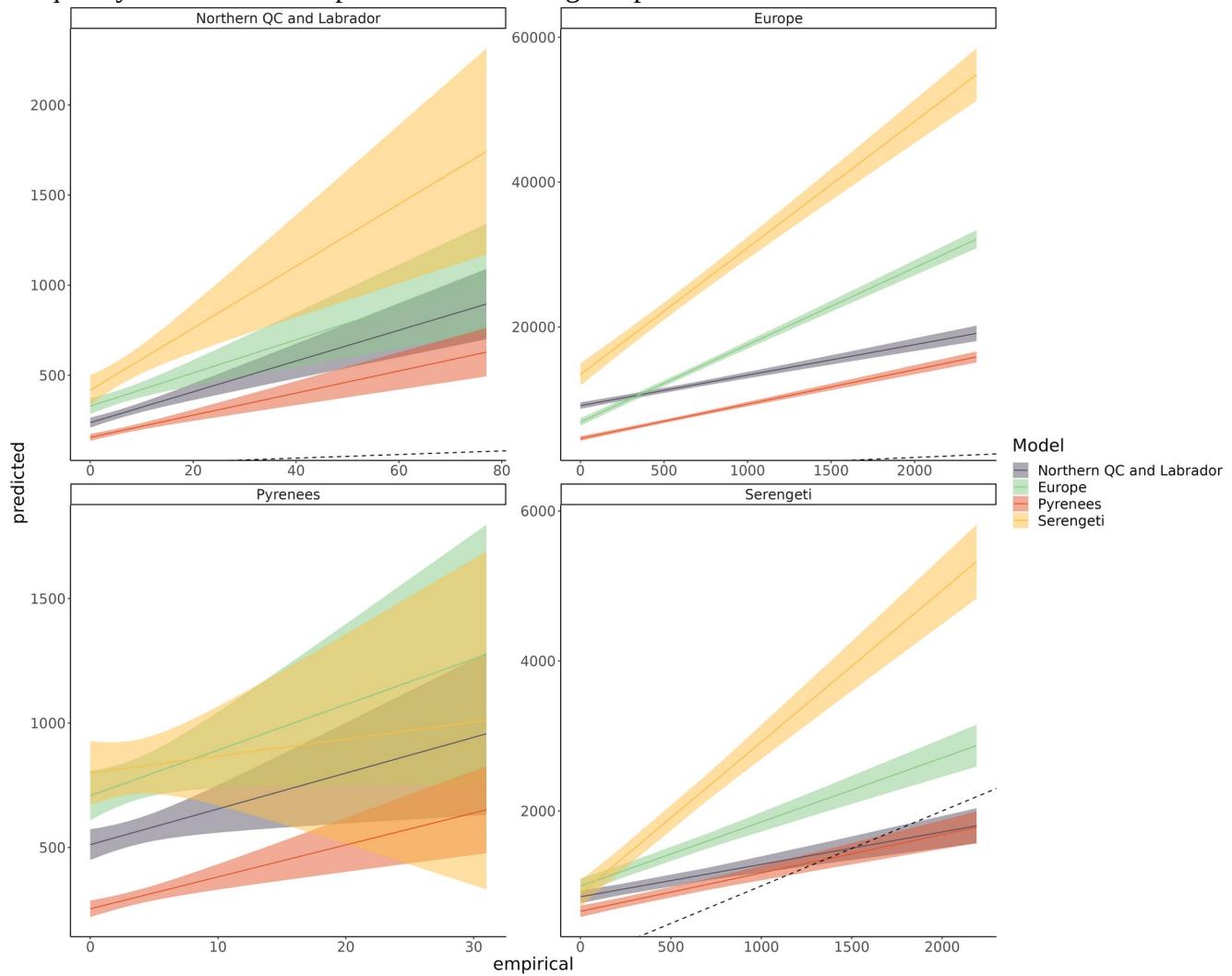
*Figure S8.13: Linear regression comparing the predicted frequency of the resource position in the linear food chain motif and empirical frequency of the resource position in the linear food chain motif of species for each model predicting every food webs. The main bias is that predicted the frequency of resource position of most species is overpredicted across the range (intercept > 0).*

### Frequency of the omnivore position in the intraguild predation motif



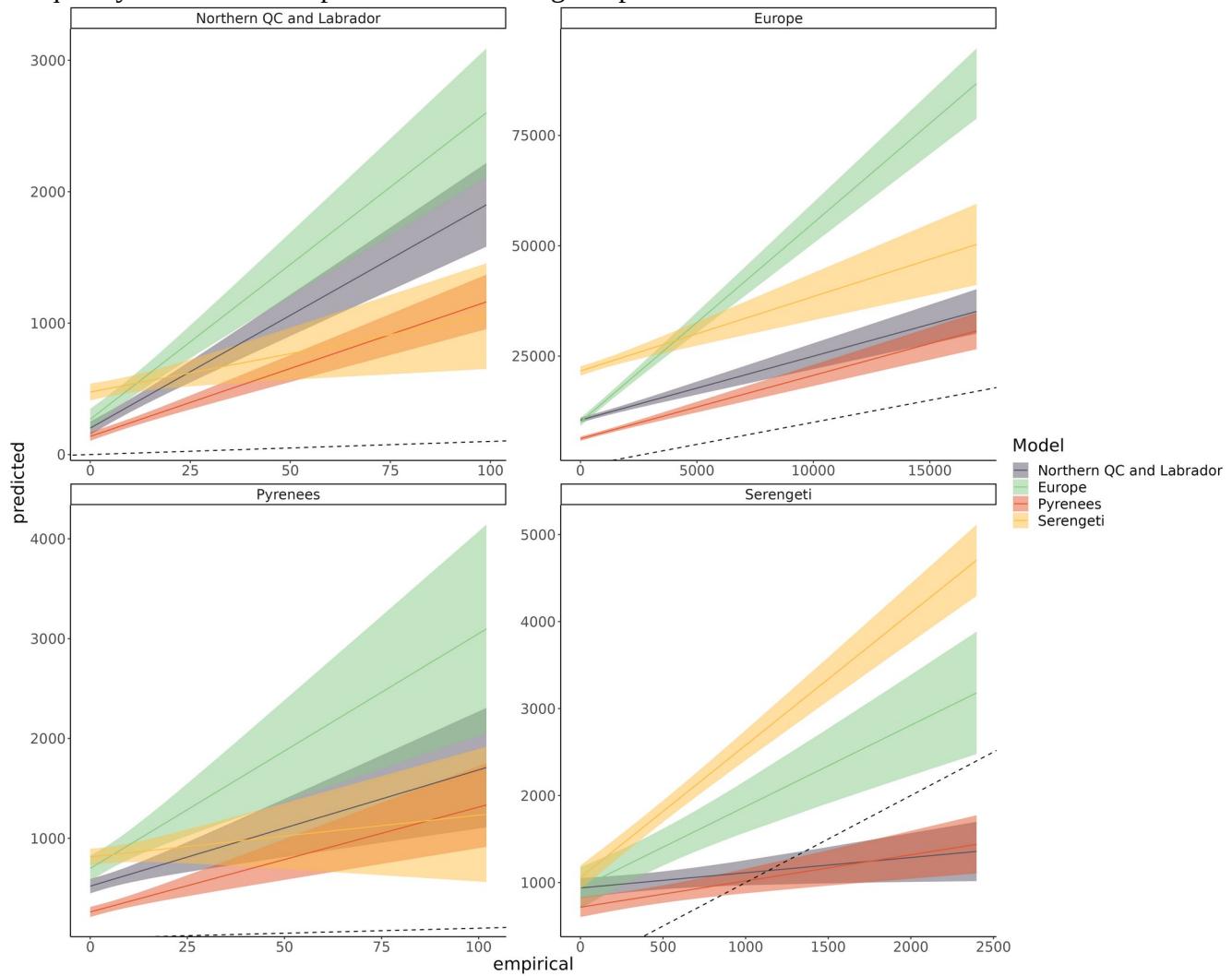
*Figure S8.14: Linear regression comparing the predicted frequency of the omnivore position in the intraguild predation motif and empirical frequency of the omnivore position in the intraguild predation motif of species for each model predicting every food webs. The main bias is that predicted the frequency of omnivore position of most species is overpredicted across the range (intercept > 0).*

### Frequency of the consumer position in the intraguild predation motif



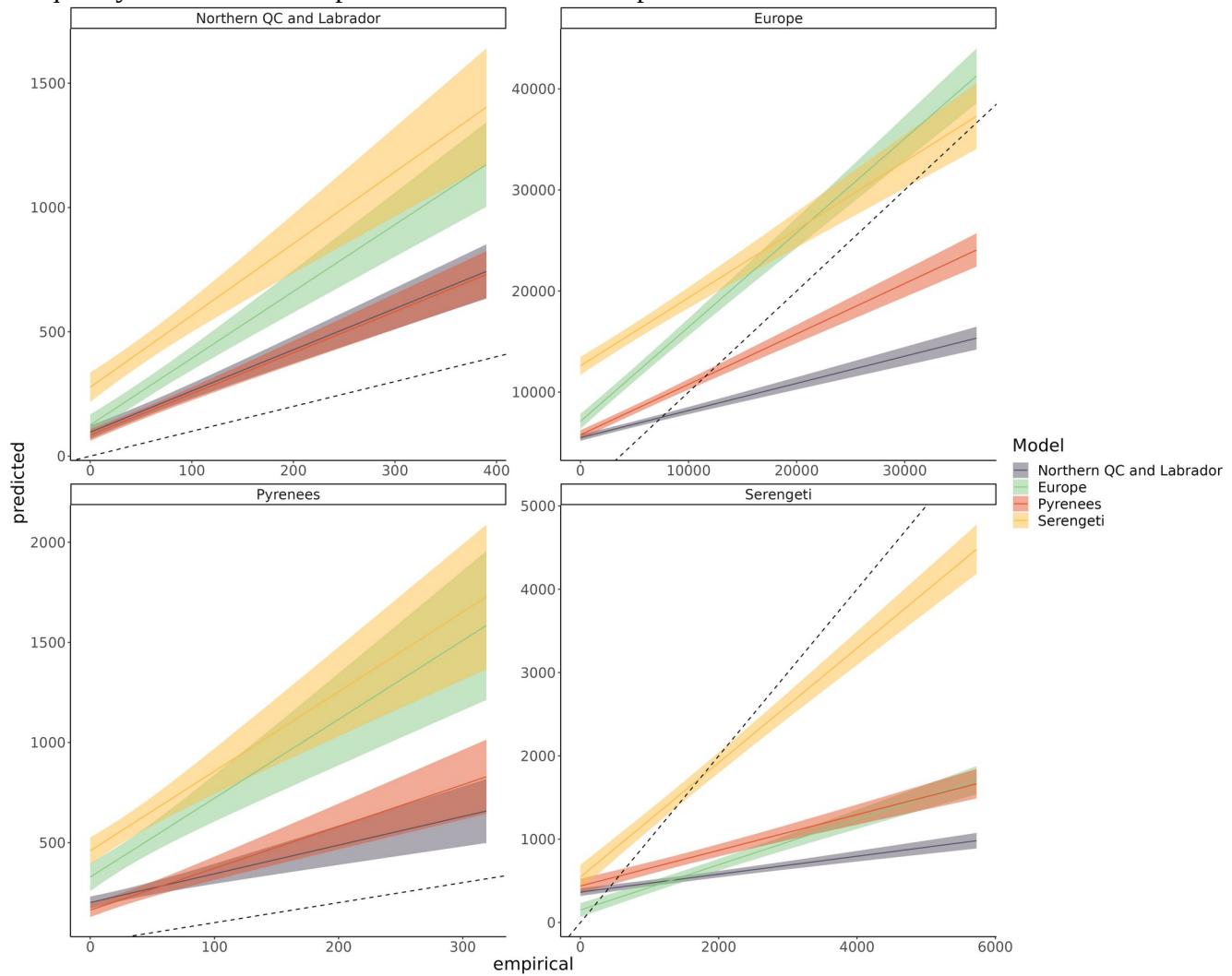
*Figure S8.15: Linear regression comparing the predicted frequency of the consumer position in the intraguild predation motif and empirical frequency of the consumer position in the intraguild predation motif of species for each model predicting every food webs. The main bias is that predicted the frequency of consumer position of most species is overpredicted across the range (intercept > 0).*

### Frequency of the resource position in the intraguild predation motif



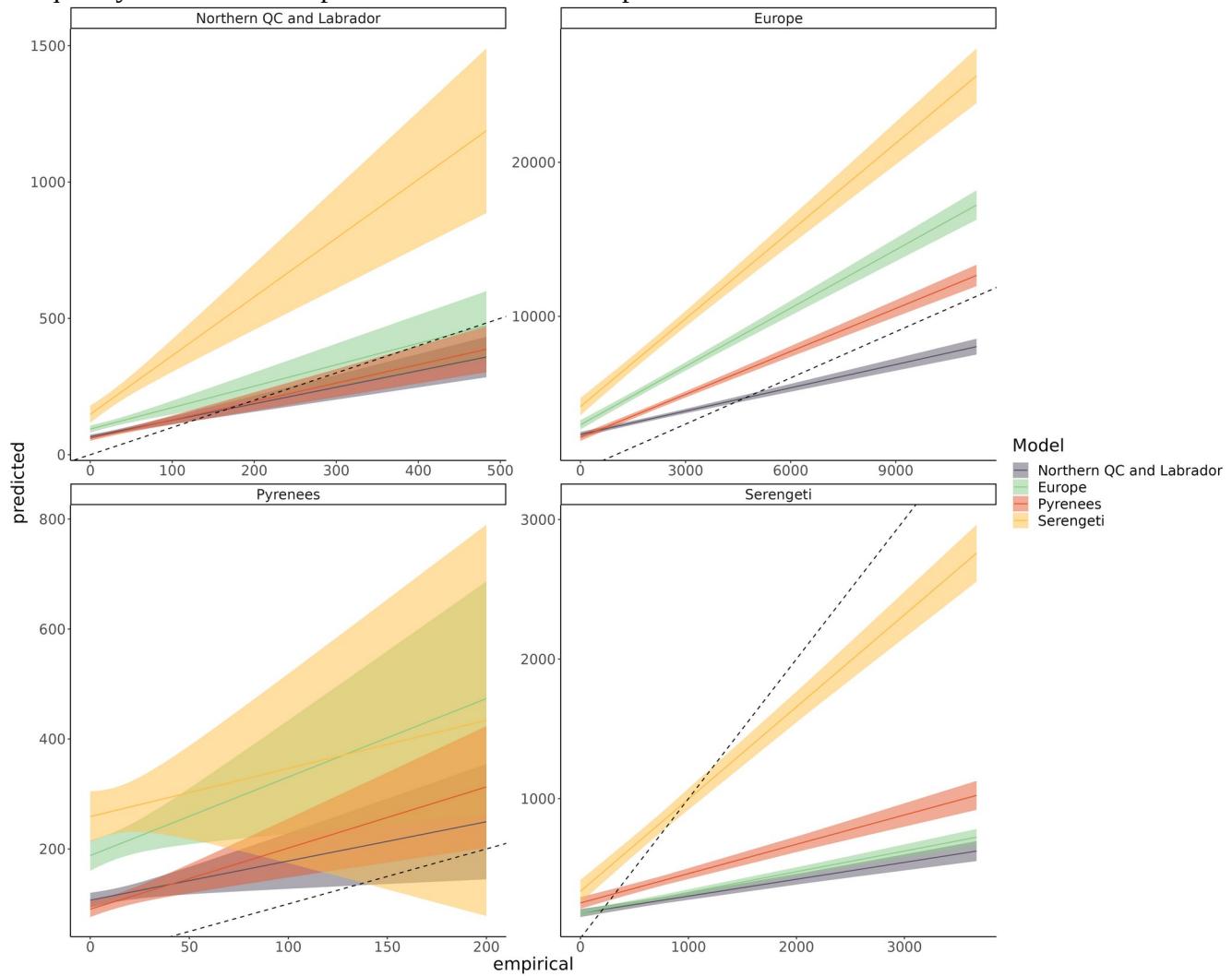
*Figure S8.16: Linear regression comparing the predicted frequency of the resource position in the intraguild predation motif and empirical frequency of the resource position in the intraguild predation motif of species for each model predicting every food webs. The main bias is that predicted the frequency of resource position of most species is overpredicted across the range (intercept > 0).*

### Frequency of the consumer position in the direct competition motif



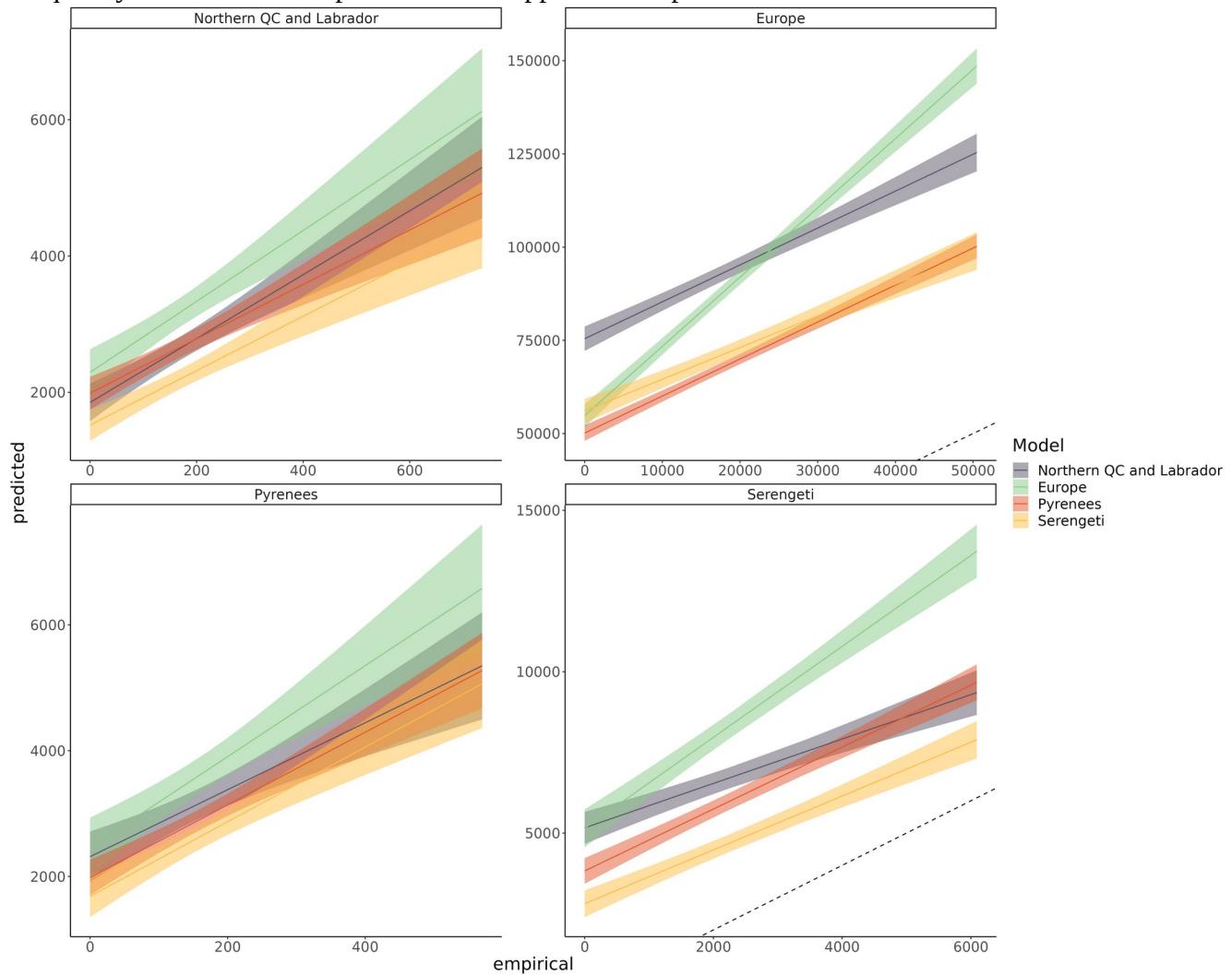
*Figure S8.17: Linear regression comparing the predicted frequency of the consumer position in the direct competition motif and empirical frequency of the consumer position in the direct competition motif of species for each model predicting every food webs. The main bias is that predicted the frequency of consumer position of most species is overpredicted across the range (intercept > 0). In the Serengeti food web, the frequency of consumer position is more homogeneous across species (slope < 1)*

### Frequency of the resource position in the direct competition motif



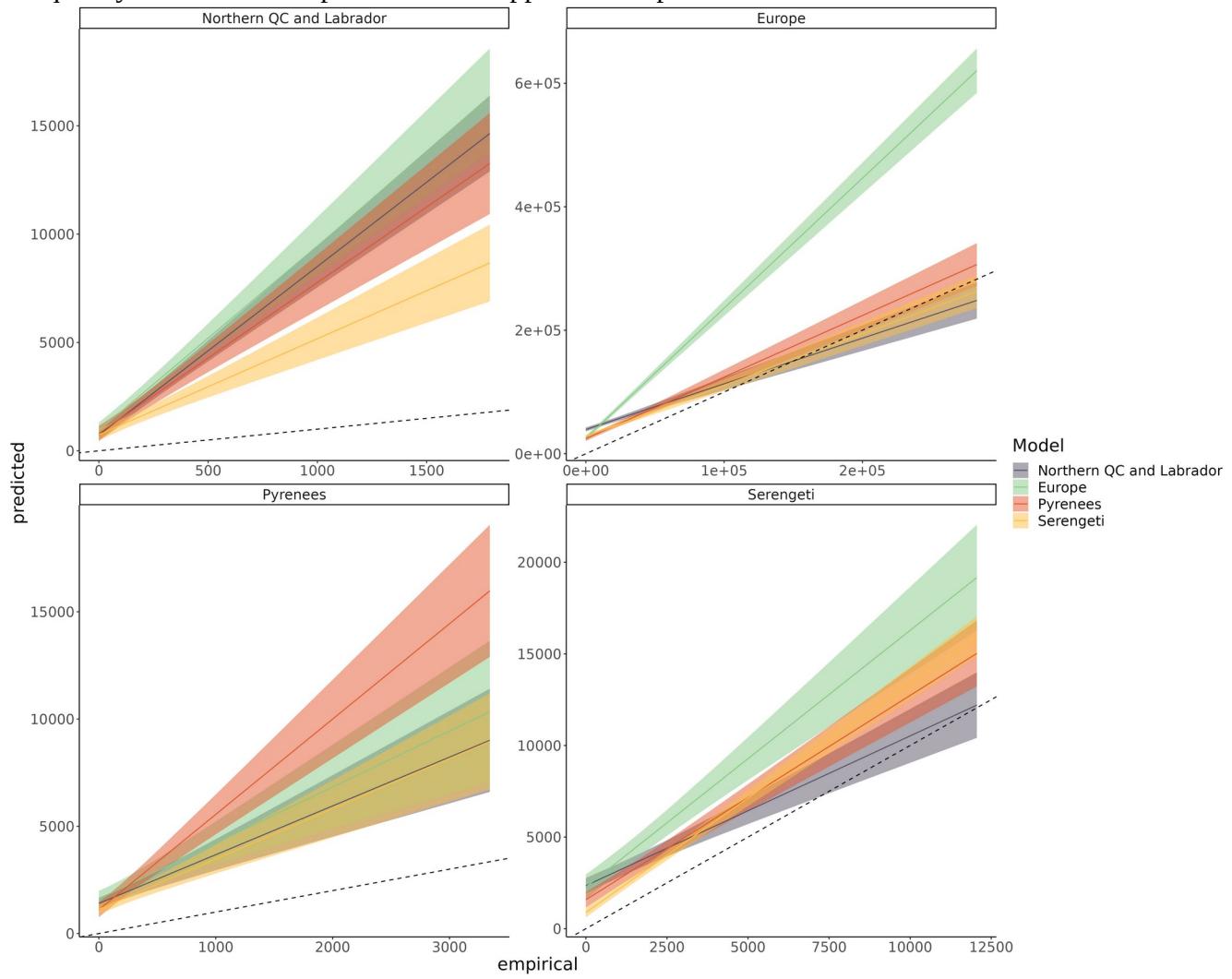
*Figure S8.18: Linear regression comparing the predicted frequency of the resource position in the direct competition motif and empirical frequency of the resource position in the direct competition motif of species for each model predicting every food webs. The main bias is that predicted the frequency of resource position of most species is overpredicted across the range (intercept > 0). In the Serengeti food web, the frequency of resource position is more homogeneous across species (slope < 1)*

### Frequency of the consumer position in the apparent competition motif



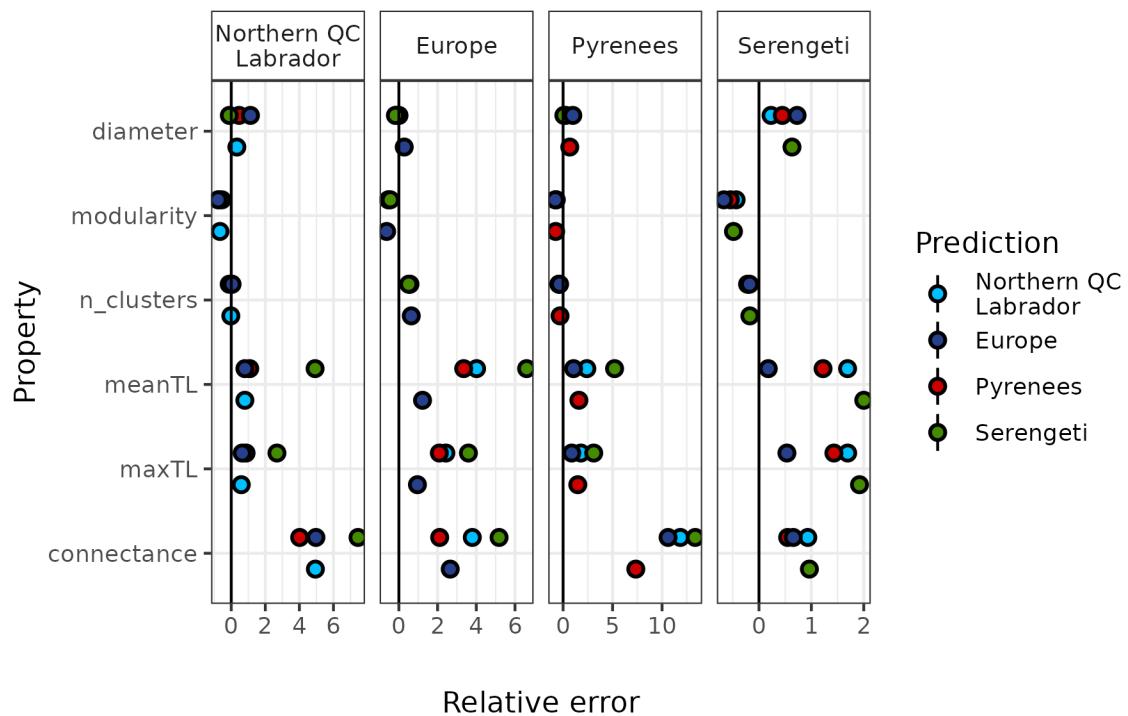
*Figure S8.19: Linear regression comparing the predicted frequency of the consumer position in the apparent competition motif and empirical frequency of the consumer position in the apparent competition motif of species for each model predicting every food webs. The main bias is that predicted the frequency of consumer position of most species is overpredicted across the range (intercept > 0).*

### Frequency of the resource position in the apparent competition motif



*Figure S8.20: Linear regression comparing the predicted frequency of the resource position in the apparent competition motif and empirical frequency of the resource position in the apparent competition motif of species for each model predicting every food webs. The main bias is that predicted the frequency of resource position of most species is overpredicted across the range (intercept > 0).*

## Appendix S9: Predicting food web properties extended results



*Figure S9.1: The relative error of the predicted food web properties. Relative error is the difference between the predicted and empirical estimates divided by the empirical estimate. From top to bottom food web diameter, modularity, number of clusters, mean trophic level, maximum trophic level, and connectance.*

## **Appendix S10: Results with alternative Serengeti food web and without the Serengeti**

### **Methods**

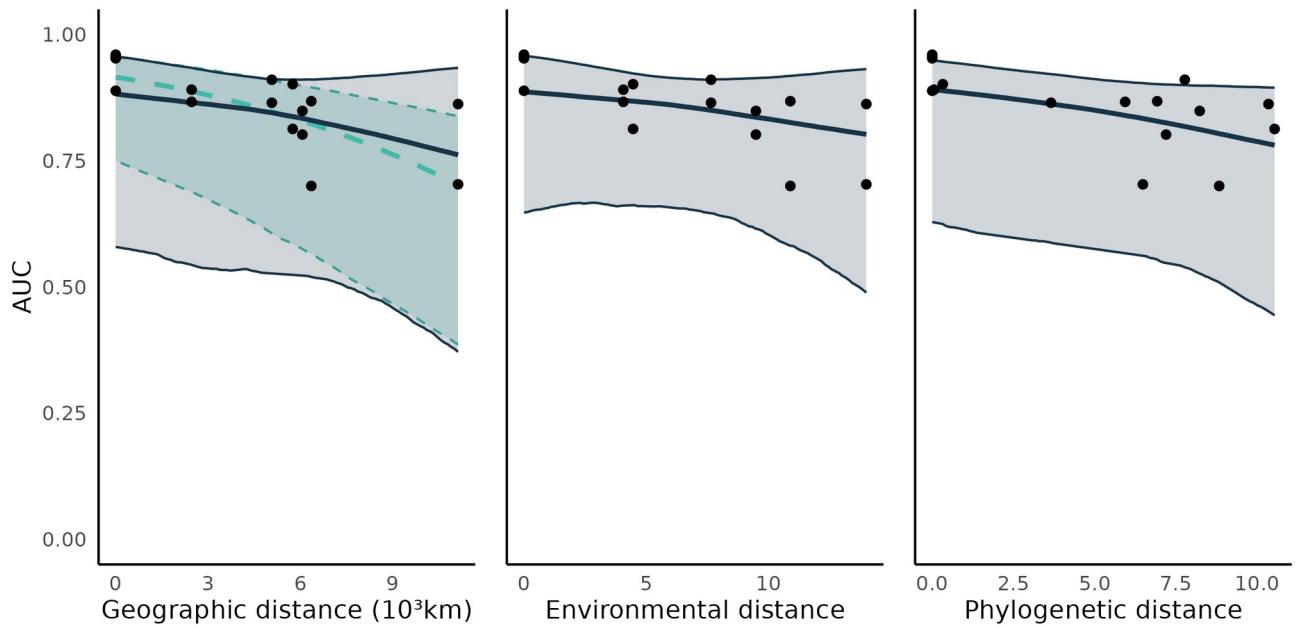
The nodes in the original Serengeti food web are originally trophic groups. In the main manuscript, we assume that all species within a trophic group share all the same prey and predators. We made this decision based on our experience: such aggregation is not that far from what is oftentimes done by experts when filling gaps in the other food webs (e.g., “all fox-like species could eat all small mammals”). To make sure this assumption does not qualitatively influence our findings, we replicated our results without including the Serengeti food web, and by keeping the nodes of the Serengeti food web as trophic groups. To get the predictors of each node, we used the mean (for continuous predictors) and the mode (for categorical predictors) across species within each trophic group, ignoring within-group trait variations.

### **Results with the Serengeti trophic groups as nodes:**

#### *Predicting species interactions*

*Table S10.1: Area under the receiver operating curve (AUC) and area under the precision-recall-gain curve (AUPRG) of each food web model predicting every food web. Values in red are better predicted using the Serengeti food web used in the main manuscript (species as nodes), whereas values in blue are better predicted using the Serengeti with trophic groups as nodes.*

Model	Food web			
	Europe	Pyrenees	North Québec and Labrador	Serengeti
AUC				
Europe	0.96	0.89	0.9	0.87
Pyrenees	0.87	0.95	0.86	0.85
North Québec And Labrador	0.81	0.91	0.95	0.86
Serengeti	0.7	0.8	0.7	0.89
AUPRG				
Europe	0.97	0.66	0.86	0.87
Pyrenees	0.9	0.98	0.95	0.92
North Québec And Labrador	0.88	0.91	0.96	0.94
Serengeti	0.72	0.79	0.81	0.88

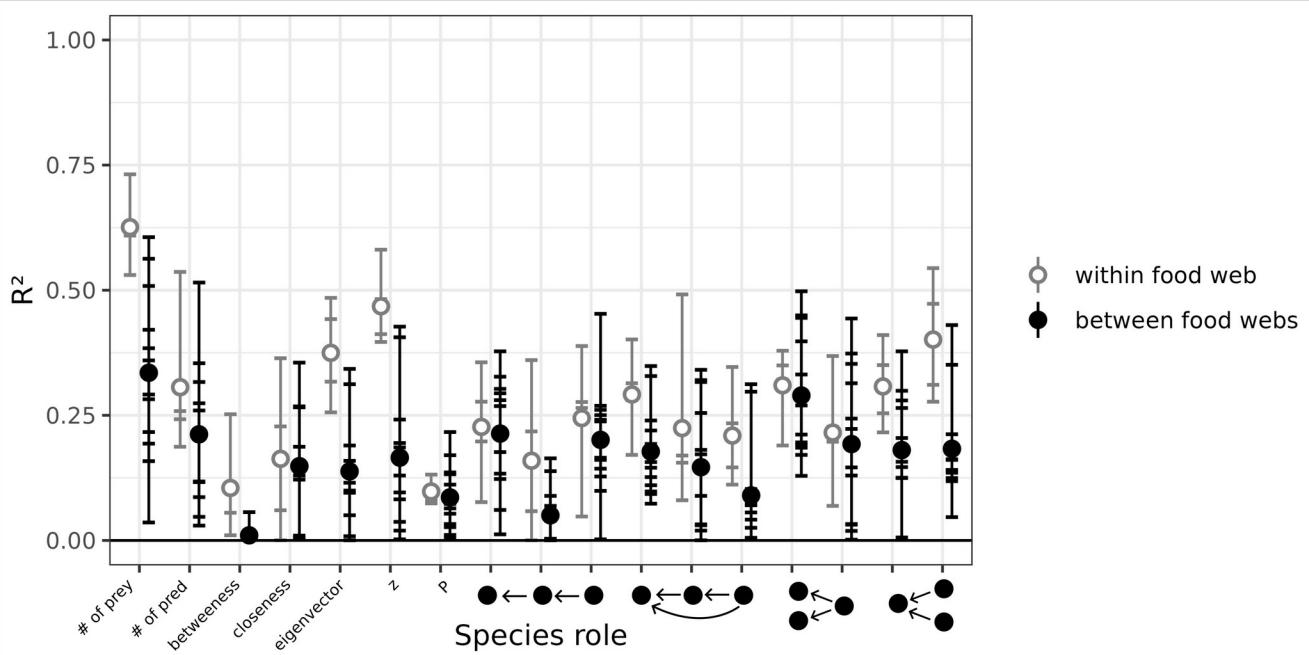


**Figure S10.1: Transferability of predictive models.** Each point is the predictive performance (AUC) of a model-food web predicted combination (16 combinations; Table S10.1). The trend lines are the median effects with their 95% credible interval constructed with the posterior predictive distribution of geographic, environmental, and phylogenetic distances on predictive performance. Shown are the total (turquoise;  $-0.49$  (95% CrI = [-0.73, -0.24])) and partial (dark blue;  $-0.25$  (95% CrI = [-0.91, 0.44])) effects of geographic distance , and the direct (controlling for geographic distance) effects of environmental ( $-0.20$  (95% CrI = [-0.90, 0.48])) and phylogenetic distances ( $-0.33$  (95% CrI = [-0.77, 0.11])).

In the alternative version of the Serengeti food web, nodes are groups of species. So, we did not replicate the results for species specific performance because we judged unreasonable to compare the predictability of interactions to phylogenetic distance and functional distance for nodes in the Serengeti to other food webs.

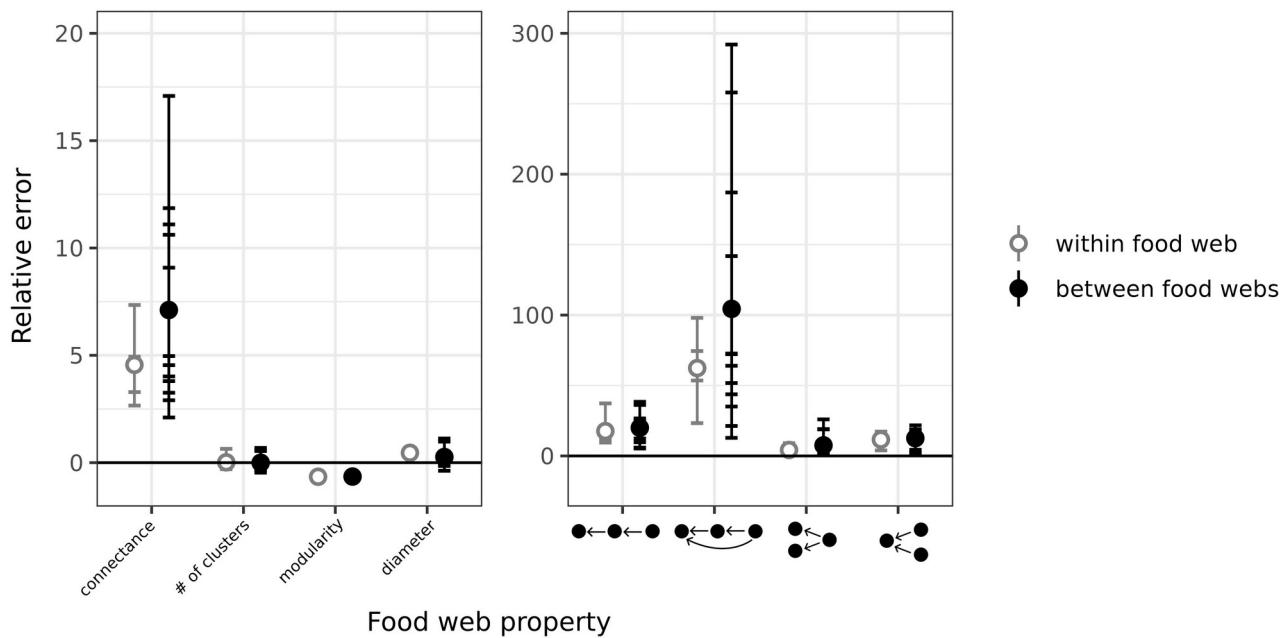
Overall, the results are similar to what we obtained using the version of the Serengeti food web we used in the main manuscript. The Serengeti food web with trophic group as nodes is better predicted by all models, but models calibrated on it predicted less well the other food webs (Table S10.1). A possible reason for this difference is that, by using a coarser species resolution, we loose small variations in the trait-interaction relationships making the interactions between trophic groups less noisy, but we loose information useful to predict species-species interactions in other food webs.

#### *Predicting species' roles*



*Figure S10.2: Performance of predictive models to predict species' role measured as the proportion of the variance in trophic positions explained by the models ( $R^2$ ). From left to right, the figure shows the number of prey, number of predators, betweenness, closeness, eigenvector centrality, within-module degree ( $z$ ), participation coefficient ( $P$ ), times the species is at a specific position in the linear chain (predator – consumer - resource), intraguild predation (omnivore – consumer - resource), direct competition (2 consumers – 1 resource), and apparent competition motifs (1 consumer – 2 resources). The short horizontal lines highlight the  $R^2$  for each role, model, and food web predicted combination. Grey open dots are the mean  $R^2$  for within food web predictions. Full black dots are the mean  $R^2$  for between food web predictions.*

## Predicting food web properties

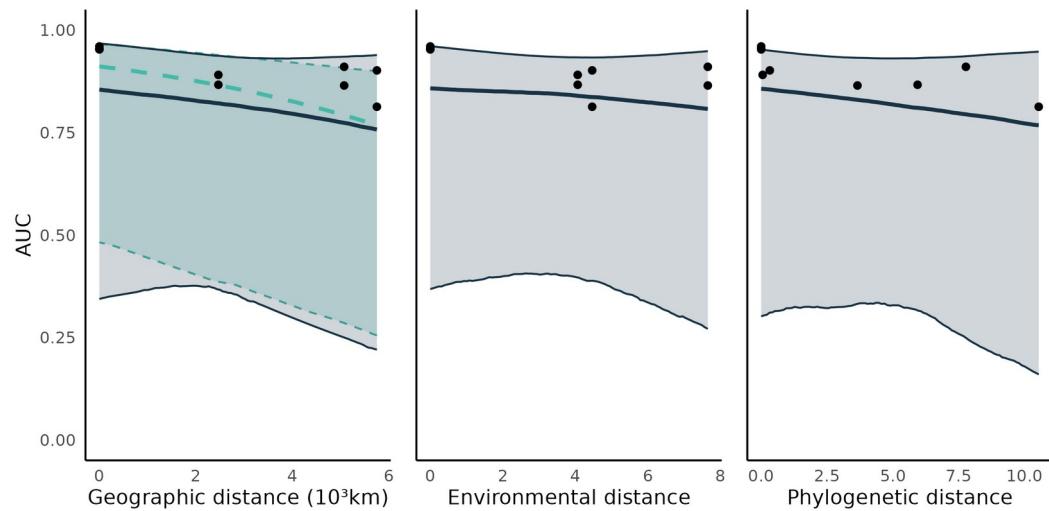


*Figure S10.3: Relative error of global food web properties predicted by the boosted regression trees. From left to right, the figure shows the relative error for connectance, number of clusters, modularity, diameter, number of linear chain (predator – consumer - resource), intraguild predation (omnivore – consumer - resource), direct competition (2 consumers – 1 resource), and apparent competition (1 consumer – 2 resources) motifs. Short horizontal lines are the relative errors for each property, model, and food web predicted combination. Grey open dots are the mean relative errors for within food web predictions and full black dots are the mean relative errors for between food web predictions.*

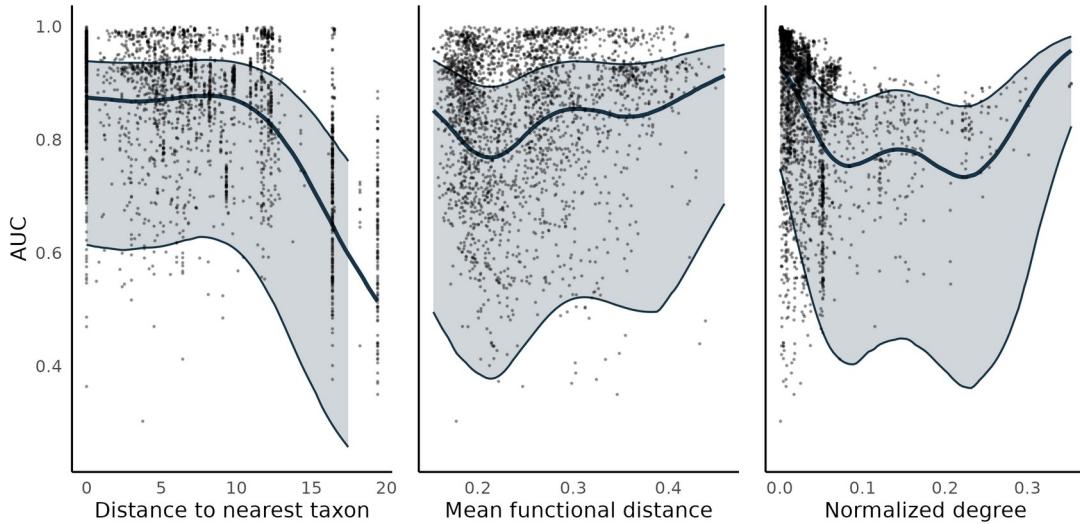
Overall, for both species' roles and food web properties, we do not find any differences in our results. Species' roles are still not well predicted, and predicted food webs are more connected and less modular. The results for trophic levels is not shown here because the European model predicted cycles in the Serengeti food web, interacting with how trophic levels are calculated.

## **Results after excluding the Serengeti food web**

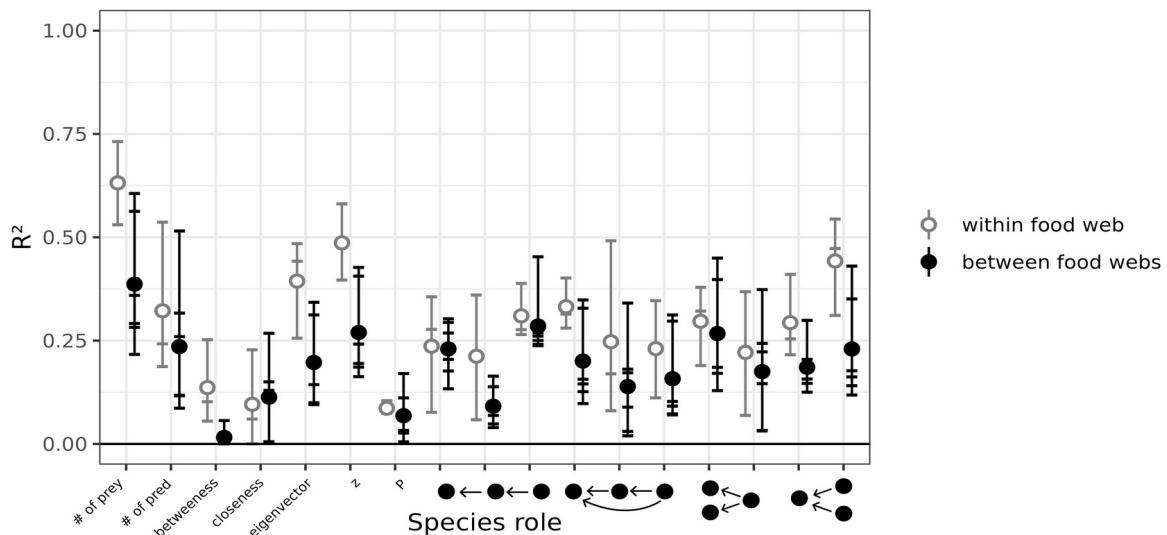
### *Predicting species interactions*



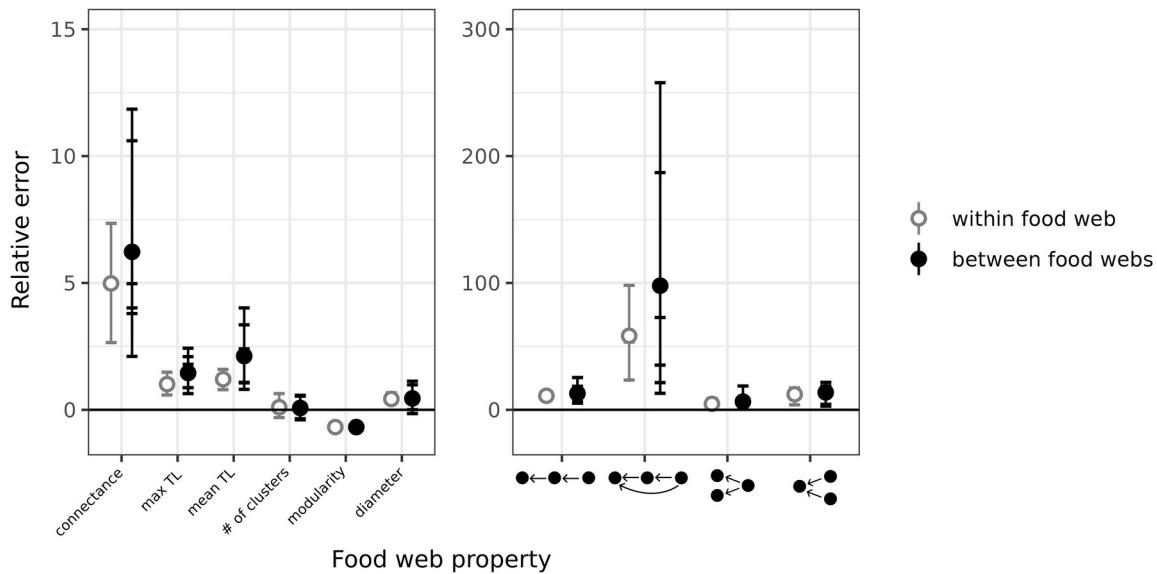
**Figure S10.4: Transferability of predictive models.** Each point is the predictive performance (AUC) of a model-food web predicted combination (12 combinations). The trend lines are the median effects with their 95% credible interval constructed with the posterior predictive distribution of geographic, environmental, and phylogenetic distances on predictive performance. Shown are the total (turquoise; -0.48 (95% CrI = [-0.90, -0.03])) and partial (dark blue; -0.28 (95% CrI = [-1.17, 0.65])) effects of geographic distance , and the direct (controlling for geographic distance) effects of environmental (-0.13 (95% CrI = [-0.94, 0.62])) and phylogenetic distances (-0.23 (95% CrI = [-1.08, 0.62])).



**Figure S10.5:** Model predictive performance for the species interactions in alternative food webs. Each point shows the model performance in predicting the prey and predators of a single species using area under the receiver operating curve (AUC) as a metric. For each panel, the thick line is the median effect while the enveloppe is the 95% credible interval on predictive performance built using the posterior predictive distribution of (a) distance to the nearest taxon, (b) mean functional pairwise distance, and (c) normalized degree on predictive performance.



**Figure S10.6:** Predicting species' role measured as the proportion of the variance in trophic positions explained by the models ( $R^2$ ). From left to right, the figure shows the number of prey, number of predators, betweenness, closeness, eigenvector centrality, within-module degree ( $z$ ), participation coefficient ( $P$ ), times the species is at a specific position in the linear chain (predator – consumer - resource), intraguild predation (omnivore – consumer - resource), direct competition (2 consumers – 1 resource), and apparent competition motifs (1 consumer – 2 resources).



*Figure S10.7: Relative error of global food web properties predicted by the boosted regression trees. From left to right, the figure shows the relative error for connectance, number of clusters, modularity, diameter, number of linear chain (predator – consumer - resource), intraguild predation (omnivore – consumer - resource), direct competition (2 consumers – 1 resource), and apparent competition (1 consumer – 2 resources) motifs.*

Overall, the results are similar to what we obtain when keeping the Serengeti food web. The effects of environmental and phylogenetic distances are slightly weaker, and credible intervals are overall larger, which is expected as we reduce sample sizes.

## Appendix S11 – References

- Allesina, S., & Pascual, M. (2009). Googling Food Webs: Can an Eigenvector Measure Species' Importance for Coextinctions? *PLOS Computational Biology*, 5(9), e1000494.  
<https://doi.org/10.1371/journal.pcbi.1000494>
- Bascompte, J., Melián, C. J., & Sala, E. (2005). Interaction strength combinations and the overfishing of a marine food web. *Proceedings of the National Academy of Sciences*, 102(15), 5443–5447.  
<https://doi.org/10.1073/pnas.0501562102>
- Berteaux, D., Ropars, P., & Casajus, N. (2018). Toundra Nunavik: Matrice des relations trophiques entre espèces du Labrador et du Québec nordique, v. 1.0 (1980-2010). *Nordicana D36*.  
<https://doi.org/10.5885/45555CE-DA1FF11FA4254703>
- Borrelli, J. J. (2015). Selection against instability: Stable subgraphs are most frequent in empirical food webs. *Oikos*, 124(12), 1583–1588. <https://doi.org/10.1111/oik.02176>
- Cirtwill, A. R., Dalla Riva, G. V., Gaiarsa, M. P., Bimler, M. D., Cagua, E. F., Coux, C., & Dehling, D. M. (2018). A review of species role concepts in food webs. *Food Webs*, 16, e00093.  
<https://doi.org/10.1016/j.fooweb.2018.e00093>
- Csardi, G., & Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal, Complex Systems*, 1695.
- Curtsdotter, A., Binzer, A., Brose, U., de Castro, F., Ebenman, B., Eklöf, A., Riede, J. O., Thierry, A., & Rall, B. C. (2011). Robustness to secondary extinctions: Comparing trait-based sequential deletions in static and dynamic food webs. *Basic and Applied Ecology*, 12(7), 571–580.  
<https://doi.org/10.1016/j.baae.2011.09.008>
- de Visser, S. N., Freymann, B. P., & Olff, H. (2011). The Serengeti food web: Empirical quantification and analysis of topological changes under increasing human impact. *Journal of Animal Ecology*, 80(2), 484–494. <https://doi.org/10.1111/j.1365-2656.2010.01787.x>

- Delmas, E., Besson, M., Brice, M.-H., Burkle, L. A., Riva, G. V. D., Fortin, M.-J., Gravel, D., Guimarães, P. R., Hembry, D. H., Newman, E. A., Olesen, J. M., Pires, M. M., Yeakel, J. D., & Poisot, T. (2019). Analysing ecological networks of species interactions. *Biological Reviews*, 94(1), 16–36. <https://doi.org/10.1111/brv.12433>
- Duffy, J. E., Cardinale, B. J., France, K. E., McIntyre, P. B., Thébaud, E., & Loreau, M. (2007). The functional role of biodiversity in ecosystems: Incorporating trophic complexity. *Ecology Letters*, 10(6), 522–538. <https://doi.org/10.1111/j.1461-0248.2007.01037.x>
- Dunne, J. A., Williams, R. J., & Martinez, N. D. (2002). Food-web structure and network theory: The role of connectance and size. *Proceedings of the National Academy of Sciences*, 99(20), 12917–12922. <https://doi.org/10.1073/pnas.192407699>
- Dyer, L. A., & Letourneau, D. (2003). Top-down and bottom-up diversity cascades in detrital vs. Living food webs. *Ecology Letters*, 6(1), 60–68. <https://doi.org/10.1046/j.1461-0248.2003.00398.x>
- Etard, A., Morrill, S., & Newbold, T. (2020). Global gaps in trait data for terrestrial vertebrates. *Global Ecology and Biogeography*, 29(12), 2143–2158. <https://doi.org/10.1111/geb.13184>
- Flach, P., & Kull, M. (2015). Precision-Recall-Gain Curves: PR Analysis Done Right. In C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, & R. Garnett (Eds.), *Advances in Neural Information Processing Systems* (Vol. 28). Curran Associates, Inc.  
[https://proceedings.neurips.cc/paper\\_files/paper/2015/file/33e8075e9970de0cfea955af4644bb2-Paper.pdf](https://proceedings.neurips.cc/paper_files/paper/2015/file/33e8075e9970de0cfea955af4644bb2-Paper.pdf)
- Freeman, L. C. (1977). A Set of Measures of Centrality Based on Betweenness. *Sociometry*, 40(1), 35–41. <https://doi.org/10.2307/3033543>
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian Data Analysis* (0 ed.). Chapman and Hall/CRC. <https://doi.org/10.1201/b16018>

- Greenwell, B., Boehmke, B., Cunningham, J., Developers, G., & Greenwell, M. B. (2019). Package ‘gbm.’ *R Package Version*, 2(5).
- Guimerà, R., & Amaral, L. A. N. (2005). Cartography of complex networks: Modules and universal roles. *Journal of Statistical Mechanics: Theory and Experiment*, 2005(02), P02001.  
<https://doi.org/10.1088/1742-5468/2005/02/P02001>
- Hijmans, R. J., Phillips, S., Leathwick, J., Elith, J., & Hijmans, M. R. J. (2017). Package ‘dismo.’ *Circles*, 9(1), 1–68.
- Jordán, F. (2009). Keystone species and food webs. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1524), 1733–1741. <https://doi.org/10.1098/rstb.2008.0335>
- Kones, J. K., Soetaert, K., Oevelen, D. van, & Owino, J. (2009). Are network indices robust indicators of food web functioning? A Monte Carlo approach. *Ecological Modelling*, 220, 370–382.  
<http://dx.doi.org/10.1016/j.ecolmodel.2008.10.012>
- Lurgi, M., López, B. C., & Montoya, J. M. (2012). Climate change impacts on body size and food web structure on mountain ecosystems. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1605), 3050–3057. <https://doi.org/10.1098/rstb.2012.0239>
- Maiorano, L., Montemaggioli, A., Ficetola, G. F., O’Connor, L., & Thuiller, W. (2020). TETRA-EU 1.0: A species-level trophic metaweb of European tetrapods. *Global Ecology and Biogeography*, geb.13138. <https://doi.org/10.1111/geb.13138>
- Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., & Alon, U. (2002). Network Motifs: Simple Building Blocks of Complex Networks. *Science*, 298(5594), 824–827.  
<https://doi.org/10.1126/science.298.5594.824>
- Saravia, L. A. (2022). *multiweb: Ecological network analysis includind multiplex networks*.  
<https://github.com/lasaravia/multiweb>

- Stekhoven, D. J. (2013). *missForest: Nonparametric missing value imputation using random forest* [Manual].
- Stekhoven, D. J., & Buehlmann, P. (2012). MissForest—Non-parametric missing value imputation for mixed-type data. *Bioinformatics (Oxford, England)*, 28(1), 112–118.
- Stouffer, D. B., & Bascompte, J. (2010). Understanding food-web persistence from local to global scales. *Ecology Letters*, 13(2), 154–161. <https://doi.org/10.1111/j.1461-0248.2009.01407.x>
- Stouffer, D. B., & Bascompte, J. (2011). Compartmentalization increases food-web persistence. *Proceedings of the National Academy of Sciences*, 108(9), 3648–3652.  
<https://doi.org/10.1073/pnas.1014353108>
- Stouffer, D. B., Sales-Pardo, M., Sirer, M. I., & Bascompte, J. (2012). Evolutionary Conservation of Species’ Roles in Food Webs. *Science*, 335(6075), 1489–1492.  
<https://doi.org/10.1126/science.1216556>
- Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., & Bürkner, P.-C. (2021). Rank-Normalization, Folding, and Localization: An Improved R<sup>̂</sup> for Assessing Convergence of MCMC (with Discussion). *Bayesian Analysis*, 16(2), 667–718. <https://doi.org/10.1214/20-BA1221>
- Vieira, M. C., & Almeida-Neto, M. (2015). A simple stochastic model for complex coextinctions in mutualistic networks: Robustness decreases with connectance. *Ecology Letters*, 18(2), 144–152.  
<https://doi.org/10.1111/ele.12394>
- Zhao, Q., Brink, P. J. V. den, Carpentier, C., Wang, Y. X. G., Rodríguez-Sánchez, P., Xu, C., Vollbrecht, S., Gillissen, F., Vollebregt, M., Wang, S., & Laender, F. D. (2019). Horizontal and vertical diversity jointly shape food web stability against small and large perturbations. *Ecology Letters*, 22(7), 1152–1162. <https://doi.org/10.1111/ele.13282>

## C. Supplementary material for chapter 3

### Consequences of seasonal geographic and niche movements for North American bird biogeography

#### Table of Contents:

	Page
<b>Appendix S1: Latitudinal and trait bias of species included in analyses</b>	<b>Page 2</b>
<b>Appendix S2: Alternative metrics of movement in geographic and climate space movements</b>	
<b>Appendix S3: Phylogenetic multilevel models</b>	<b>Page</b>
<b>Appendix S4: Model convergence</b>	<b>Page</b>

## **Appendix S1: Latitudinal and trait bias of species included in analyses**

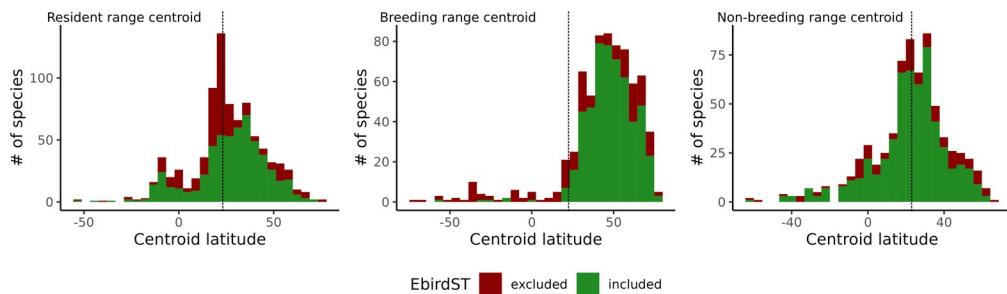
### **Description:**

In this appendix, we explore the bias across latitude and traits of North American bird species included in our analyses ( $N = 546$ ) compared to the entire assemblage of North American bird species ( $N = 824$ ) in the Avibase checklist (Lepage *et al.* 2014). We investigate these biases to check if the missing species could influence our analyses of the latitudinal diversity gradient and the trend between body mass and latitude (Bergmann's rule) across seasons.

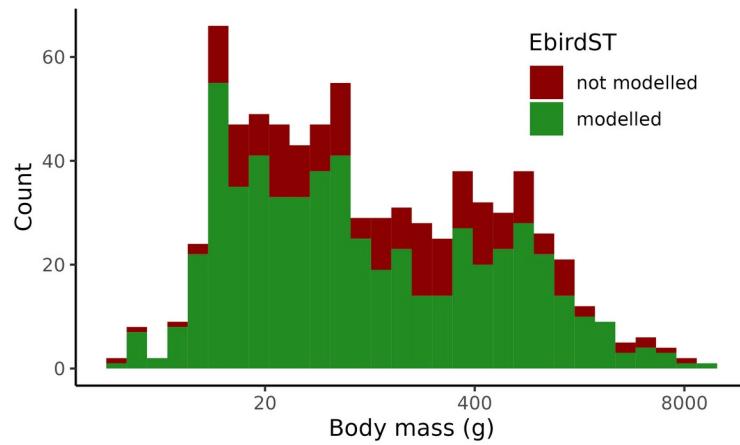
To explore the latitudinal bias, we downloaded the BirdLife range maps (BirdLife International and Handbook of the Birds of the World, 2022) for the 824 bird species, and calculated the centroid of the resident (area where the species is thought very likely to be resident throughout the year), breeding (area where the species is thought very likely to occur regularly during the breeding season), and non-breeding range (area where the species is thought very likely to occur regularly during the winter season) of each species. We then compared the centroids of species included and the ones excluded from our analyses.

To explore the trait bias, we extracted the body mass, hand-wing index of all species of the Avibase checklist also present in AVONET. We then compared the distribution of traits for species included and the ones excluded from our analyses.

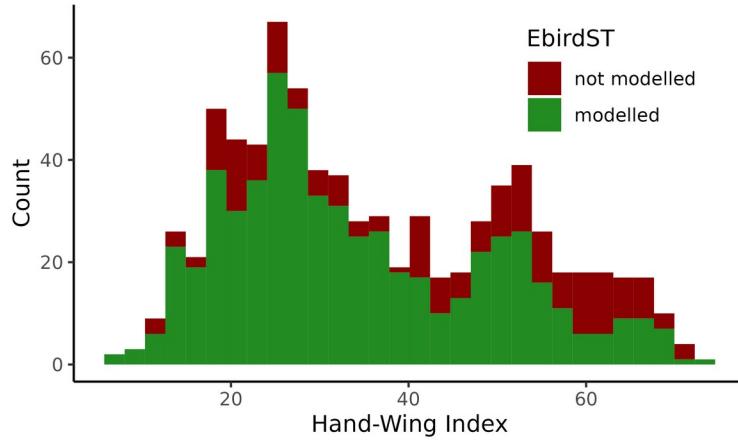
## Results:



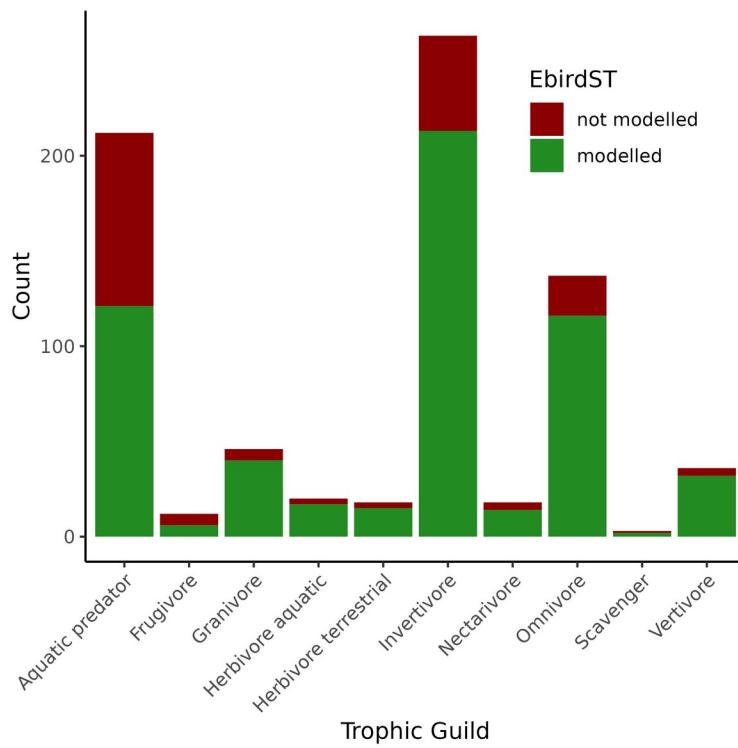
**Figure S18.1: Coverage of the North American bird assemblage across latitudes by the study.**  
The latitude of each bird range is determined by the centroid of the Birdlife range. The left panel shows the centroid of the resident range, the central panel shows the centroid of the breeding range, and the right panel shows the centroid of the non-breeding range of each species. The vertical dashed line at latitude 24° is the Southern limit of our study area (Continental United States and Canada).



**Figure S1.2: Comparison between the distribution of body mass for species included and excluded from our analyses.**



**Figure S1.3:** Comparison between the distribution of hand-wind index for species included and excluded from our analyses.



**Figure S1.4:** Comparison between the distribution of species across guilds between species included and excluded from our analyses.

### **Brief interpretation:**

Overall, within our study area, we do not detect a strong bias in species coverage across latitudes. Most species breeding in south of the Continental United States are excluded from our analyses, which is what is expected, given they are not the focus of the study. This result indicates that the sampling bias is unlikely to influence our results on the seasonal variation of ecogeographical rules. For traits, it appears there is some bias for hand-wing index with missing species having slightly larger hand-wing indices on average than modelled species. Given that the entire range of values is well represented, and we did not look at the relationship between hand-wing index across space, we are confident that our results are robust to the incomplete sampling. All guilds are well represented in our analyses. Not surprisingly, we are missing relatively more aquatic predators, given we focused on bird communities on continental United States and Canada.

### **References:**

Lepage, D., Vaidya, G., & Guralnick, R. (2014). Avibase – a database system for managing and organizing taxonomic concepts. *ZooKeys*, 420, 117–135.

<https://doi.org/10.3897/zookeys.420.7089>

BirdLife International and Handbook of the Birds of the World (2022) Bird species distribution maps of the world. Version 2022.1. Available at <http://datazone.birdlife.org/species/requestdis>.

## **Appendix S2: Alternative metrics of movement in geographic and climate space movements**

### **Description:**

In this appendix, we explore two alternative approaches to measure the movement of species in geographic (axes are latitude and longitude) and climate space (axes are scaled temperature and precipitation). The first approach is identical to the approach presented in the main text, but without the temporal smoothing ('*raw weekly movement approach*'). In this approach, we calculate amount of movement of a species by summing the distances between centroids over consecutive weeks in a 2D space where axes are scaled temperature and precipitation. With this approach, we aim to explore the effect of using the moving window average smoothing (e.g., centroid for week 1 is the average of the raw centroids of weeks 51, 52, 1, 2, and 3) on our findings.

In the second approach, we calculate the amount of movement of a species as the distance between the centroid of species during breeding and wintering seasons ('*seasonal centroids approach*'). To determine the centroid of species for these seasons, we averaged the centroids of species of weeks within their breeding and nonbreeding season. We used the start and end of each season and species as defined by the eBird Status and Trends data. This approach is more similar to many earlier work who compared the climate within the breeding and wintering IUCN ranges of species, and ignore movement at finer temporal scale.

Here, we replicate the results of the relationship between movements in geographic and climatic space (figure 1), and the relationship between traits and movements (figure 2; table 1). We additionally demonstrate the effect of using the moving window smoothing on the centroid time series for a species.

## Results:

### Raw weekly movement approach

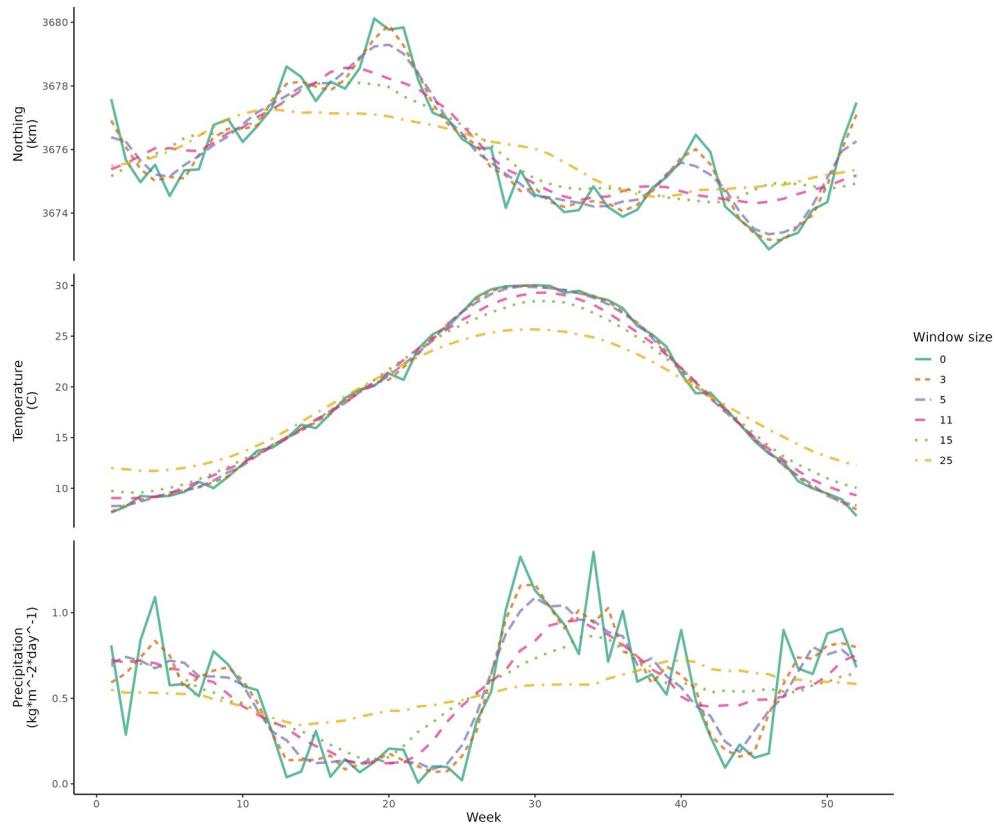
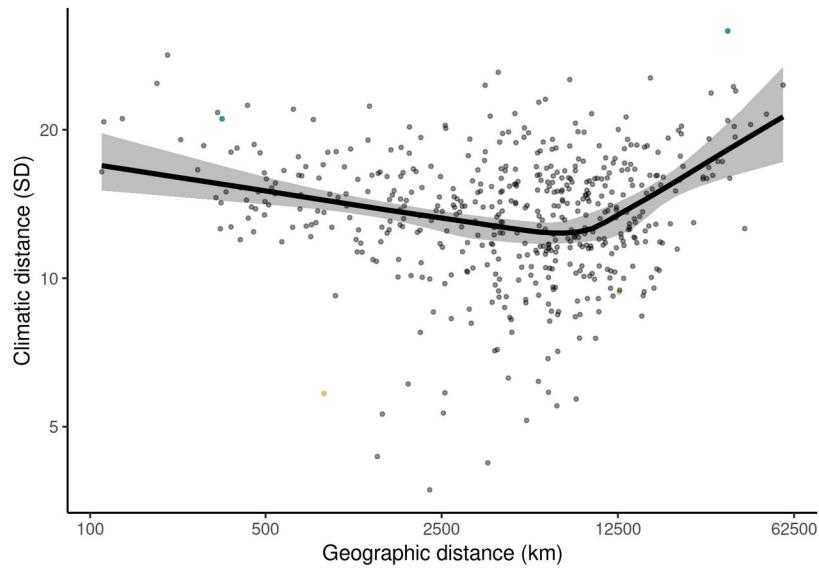
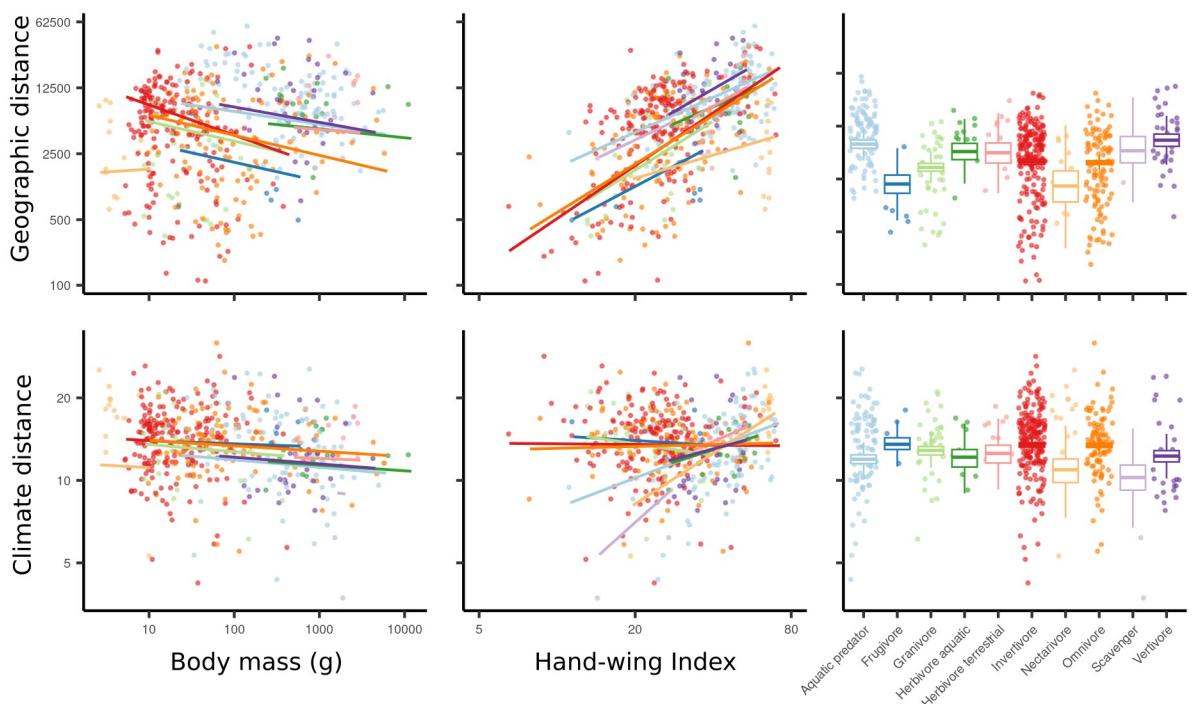


Figure S2.1: Example of the effect of the temporal smoothing on weekly geographic and climatic centroid. The figure show the Northing of geographic centroids (upper panel), mean Temperature of climate centroids (middle panel), and Precipitation rate (lower panel) of Albert's Towhee (*Melozone aberti*). We show the raw centroids (green), and five different moving window sizes.

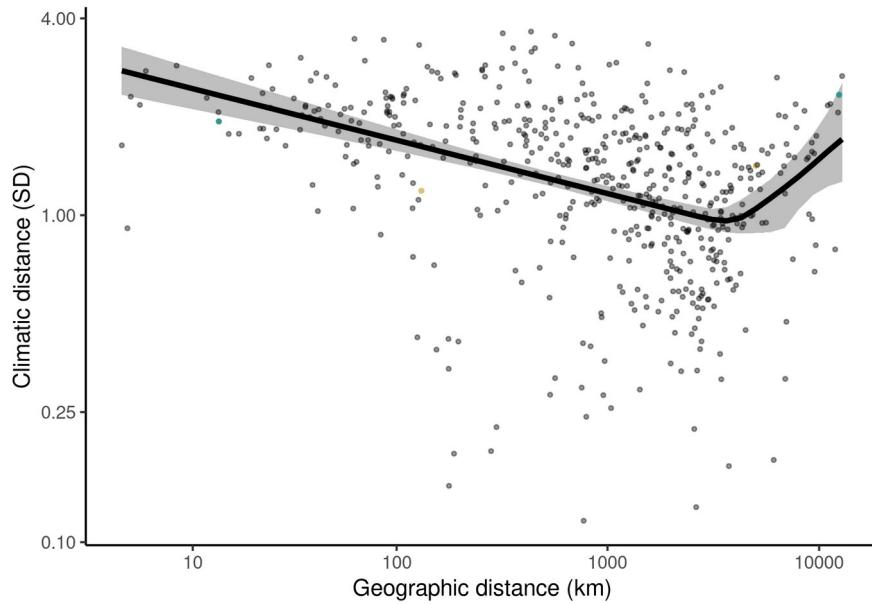


*Figure S2.2: Relationship between geographic distance and climatic distance using the raw (not smoothed) geographic and climatic centroids. Each point is the sum of distances between the centroids of consecutive weeks in climatic and geographic space weighted by abundance. The line is the mean effects while the envelop is the 95% credible interval built using the posterior predictive distribution of the piecewise model. We highlighted an example of resident ‘niche-switcher’ (Rufous-winged Sparrow), resident ‘niche-tracker’ (Great Kiskadee), migrant ‘niche-switcher’ (Buff-breasted Sandpiper), and a migrant ‘niche-tracker’ (Eastern Wood-Pewee). Turquoise points are the highlighted ‘niche-switchers’, and golden points the highlighted ‘niche-trackers’.*

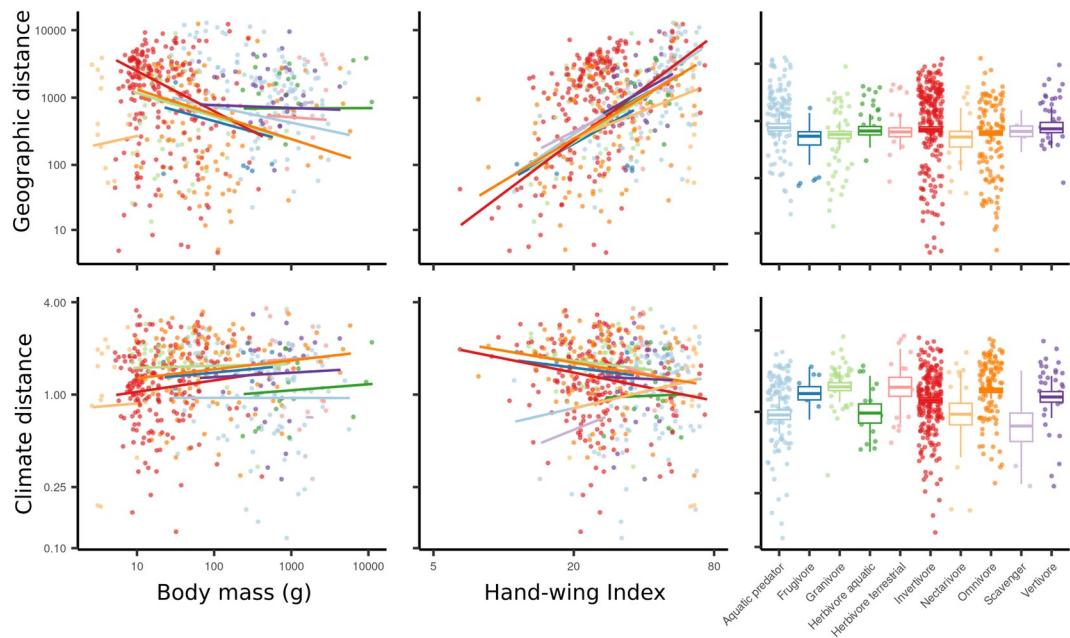


**Figure S2.3: Relationship between migratory behaviour and species traits after controlling for species phylogenetic relatedness.** Mean effect of body mass (left panels), hand-wing index (centre panels), and trophic guilds (right panels) on geographic distance (upper panels) and climatic distance (lower panels) travel by birds. Each point is a species, and the lines and boxplot are the effects determined by the posterior predictive distribution. For all traits, the figure show the direct effect of the trait (after controlling for other traits).

Seasonal centroids approach



**Figure S2.4:** Relationship between geographic distance and climatic distance using breeding and non breeding geographic and climatic centroids. The line is the mean effects while the envelop is the 95% credible interval built using the posterior predictive distribution of the piecewise model. We highlighted an example of resident ‘niche-switcher’ (Rufous-winged Sparrow), resident ‘niche-tracker’ (Great Kiskadee), migrant ‘niche-switcher’ (Buff-breasted Sandpiper), and a migrant ‘niche-tracker’ (Eastern Wood-Pewee). Turquoise points are the highlighted ‘niche-switchers’, and golden points the highlighted ‘niche-trackers’.



**Figure S2.5:** Relationship between migratory behaviour and species traits after for controlling for species phylogenetic relatedness. Mean effect of body mass (left panels), hand-wing index (centre panels), and trophic guilds (right panels) on geographic distance (upper panels) and climatic distance (lower panels) travel by birds. Each points is a species, and the lines and boxplot are the effects determined by the posterior predictive distribution. For all traits, the figure show the direct effect of the trait (after controlling for other traits).

### **Brief interpretation:**

The temporal smoothing of centroid time series helps to remove some of the noise in the data (Figure S2.1). The values of the northing of the geographic centroid and the precipitation of the climatic centroid sometimes vary widely between two weeks. Temperature seems to be less noisy across weeks. For our analysis we chose a window size of 5 weeks (e.g., Northing of the centroid for week 3 is the average northing for the centroids 1, 2, 3, 4, and 5), which, based on visual evaluation, balance the reduction of noise while keeping enough inter-week movement.

Both the ‘raw weekly centroid’ and the ‘seasonal centroid’ approaches give effects that are qualitatively similar to the effects found using the smoothed weekly centroid (reported in the main manuscript). For the relationship between movement in climatic and geographic spaces (Figure S2.2 and S2.4), species travelling more in geographic space tend to move less within climatic space, except for long-distance migrants who tend to travel more in climatic space with increasing geographic distances. The relationship between traits and movement in both spaces are also similar (Figure S2.3 and S2.5).

We do find detect some small differences. First, using the seasonal centroids, the negative effect of geographic distance on climatic distance seem to hold for more species compared to the weekly approaches. By averaging over entire seasons, we loose intra-season movements. Long-distance migrant are probably moving in climatic space more within season as they tend to breed in more seasonal northern ecosystems. Second, we see more differences across guilds for geographic movement when using weekly centroids, but less differences for climatic movements. Interestingly, when using centroids, we do find the increased climatic distances travelled by species having a more flexible diet (i.e., granivores, omnivores). However, these differences are small and probably caused by intra-seasonal movements. These analyses highlight the potential difference of intraseasonal and

interseasonal movements among species. More research would be needed to understand the effect of traits in these movements at different temporal scales.s

## Appendix S3: Phylogenetic multilevel models

### Description:

In this appendix, we test whether taking into account the fact that species are not fully independent due to phylogenetic relatedness change our results. To do so, we replicate the “Climate tracking” results from the main manuscript, but controlling for phylogenetic relatedness. We fitted the same set of models but adding a varying intercept over species constrained by a covariance matrix built from 100 posterior trees from BirdTree (Jetz et al., 2012).

Like for the main manuscript, Model parameters were estimated with Hamiltonian Monte Carlo (Neal, 2011). We ran four chains each with 1000 warm-up iterations, followed by 1000 iterations for inference. We conducted the analyses using Stan (Carpenter et al., 2017) through the package brms in R (Bürkner, 2018).

Here is the list of models, associated brms formula, and output:

### Model 1: Piecewise model of climate distance as a function of geographic distance

```
log(clim_centroid.dist) ~ b0 + b1 * (log1p(geo.dist)-omega) * step(omega -  
log1p(geo.dist)) + b2 * (log1p(geo.dist) - omega) * step(log1p(geo.dist) -  
omega), b0 ~ 1 + (1|gr(phylo, cov = A)), b1 + b2 + alpha ~ 1, nlf(omega ~  
inv_logit(alpha) * 10)
```

Variables and coefficient:

- clim\_centroid.dist is the climatic distance between the breeding and wintering climatic centroid in standard deviation unit
- geo.dist is the grand circle distance between the breeding and winter geographic centroid in km.
- omega is the geographic distance of the breaking point
- b0 is the common intercept
- b1 is the slope before the breaking point

- b2 is the slope after the breaking point
- phylo is the list of species name
- A is the covariance matrix

**Model 2: Linear model of geographic distance as a function of body mass, hand-wing index, and trophic guild**

$$\text{log1p(geo.dist)} \sim \text{logBM_sc} + \text{logHWI_sc} + (1|\text{gr(phylo, cov = A)})$$

Variables:

- geo.dist is the grand circle distance between the breeding and winter geographic centroid in km.
- logHWI\_sc: Scaled and log-transformed hand-wing index
- logBM\_sc: Scaled and log-transformed body mass
- Trophic.Niche: Trophic guild
- phylo is the list of species name
- A is the covariance matrix

**Model 3: Linear model of geographic distance as a function of body mass**

$$\text{log1p(geo.dist)} \sim \text{logBM_sc} + (1|\text{gr(phylo, cov = A)})$$

Variables:

- geo.dist is the grand circle distance between the breeding and winter geographic centroid in km.
- logBM\_sc: Scaled and log-transformed body mass
- phylo is the list of species name
- A is the covariance matrix

**Model 4: Linear model of geographic distance as a function of hand-wing index,**

$$\text{log1p(geo.dist)} \sim \text{logHWI_sc} + (1|\text{gr(phylo, cov = A)})$$

Variables:

- geo.dist is the grand circle distance between the breeding and winter geographic centroid in km.
- logHWI\_sc: Scaled and log-transformed hand-wing index
- phylo is the list of species name
- A is the covariance matrix

#### **Model 5: Linear model of geographic distance as a function of trophic guild**

$$\text{log1p}(\text{geo.dist}) \sim (1|\text{Trophic.Niche}) + (1|\text{gr}(\text{phylo}, \text{cov} = \text{A}))$$

Variables:

- geo.dist is the grand circle distance between the breeding and winter geographic centroid in km.
- Trophic.Niche: Trophic guild
- phylo is the list of species name
- A is the covariance matrix

#### **Model 6: Linear model of climatic distance as a function of body mass, hand-wing index, and trophic guild**

$$\text{log}(\text{clim\_centroid.dist}) \sim \text{logBM\_sc} + \text{logHWI\_sc} + (1+\text{logBM\_sc}+\text{logHWI\_sc}|\\ \text{Trophic.Niche}) + (1|\text{gr}(\text{phylo}, \text{cov} = \text{A}))$$

Variables:

- clim\_centroid.dist is the climatic distance between the breeding and wintering climatic centroid in standard deviation unit
- logHWI\_sc: Scaled and log-transformed hand-wing index
- logBM\_sc: Scaled and log-transformed body mass
- Trophic.Niche: Trophic guild
- phylo is the list of species name
- A is the covariance matrix

#### **Model 7: Linear model of climatic distance as a function of body mass<**

$$\log(\text{clim\_centroid.dist}) \sim \log\text{BM\_sc} + (1|\text{gr}(\text{phylo}, \text{cov} = \text{A}))$$

Variables:

- clim\_centroid.dist is the climatic distance between the breeding and wintering climatic centroid in standard deviation unit
- logBM\_sc: Scaled and log-transformed body mass
- phylo is the list of species name
- A is the covariance matrix

#### **Model 8: Linear model of climatic distance as a function of hand-wing index**

$$\log(\text{clim\_centroid.dist}) \sim \log\text{HWI\_sc} + (1|\text{gr}(\text{phylo}, \text{cov} = \text{A}))$$

Variables:

- clim\_centroid.dist is the climatic distance between the breeding and wintering climatic centroid in standard deviation unit
- logHWI\_sc: Scaled and log-transformed hand-wing index
- phylo is the list of species name
- A is the covariance matrix

#### **Model 9: Linear model of climatic distance as a function of trophic guild**

$$\log(\text{clim\_centroid.dist}) \sim (1|\text{Trophic.Niche}) + (1|\text{gr}(\text{phylo}, \text{cov} = \text{A}))$$

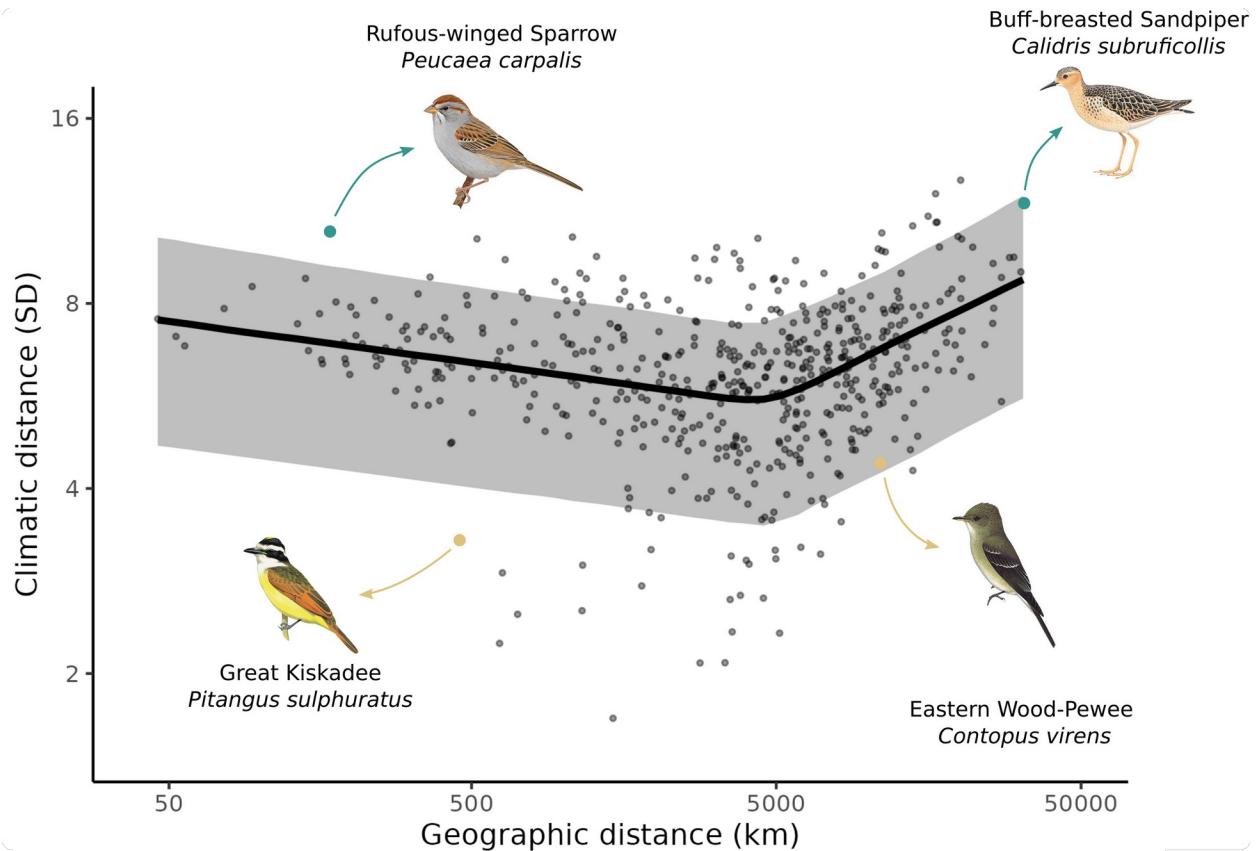
Variables:

- clim\_centroid.dist is the climatic distance between the breeding and wintering climatic centroid in standard deviation unit
- Trophic.Niche: Trophic guild
- phylo is the list of species name
- A is the covariance matrix

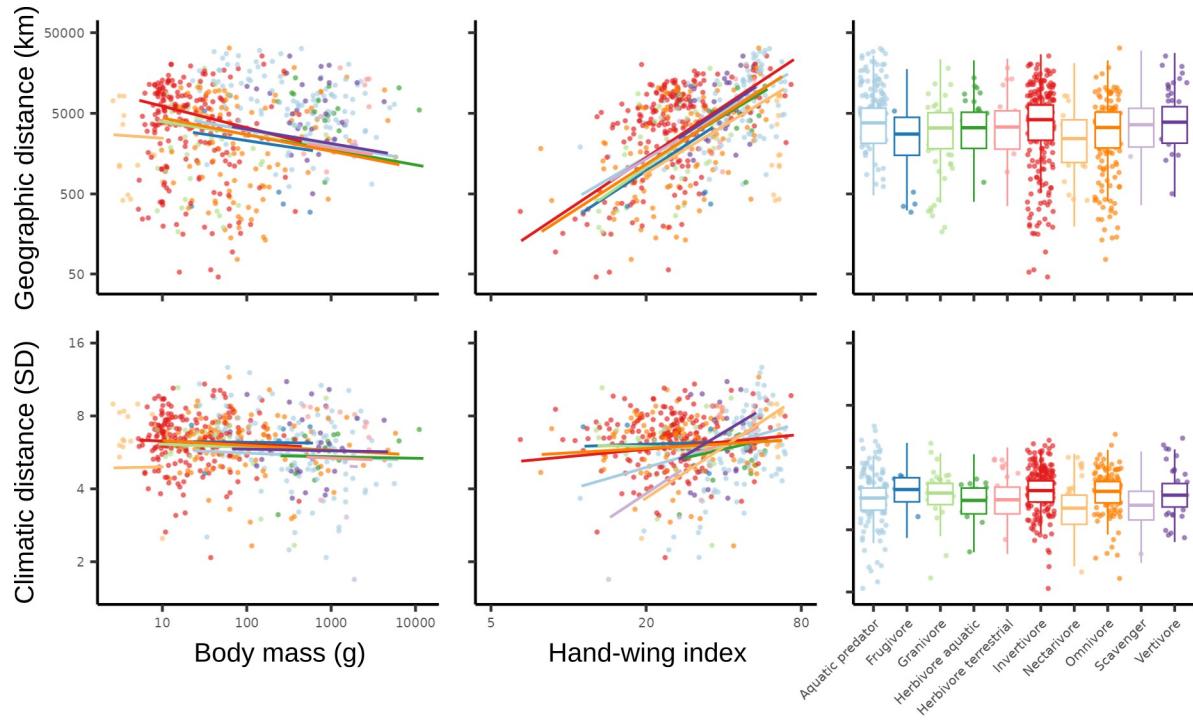
**Results:**

*Table S2.1: Effect of each traits on migration distance (A) and climatic distance (B). Univariate regression columns show the total effect of each traits (not controlling for other traits), whereas multivariate regression columns show the direct effect of each traits (controlling for other traits). The values are the mean effect and the values in square bracket the 95% credible interval from the posterior predictive distribution of the models.*

(A)	Migration distance			
	Univariate regression	Multivariate regression		
		Effect	Intercept	Body mass
<b>Trophic guild</b>				
Aquatic predator	7.95 [6.90, 8.98]	7.99 [7.17, 8.84]	-0.34 [-0.59, -0.08]	0.84 [0.54, 1.05]
Frugivore	7.46 [6.09, 8.64]	7.68 [6.57, 8.68]	-0.29 [-0.72, 0.27]	0.94 [1.56, 1.36]
Granivore	7.70 [6.58, 8.73]	7.85 [7.00, 8.72]	-0.29 [-0.63, 0.10]	1.00 [0.79, 1.34]
Herbivore aquatic	7.78 [6.66, 8.87]	7.86 [6.92, 8.81]	-0.34 [-0.69, 0.05]	0.86 [0.38, 1.18]
Herbivore terrestrial	7.88 [6.75, 9.04]	7.87 [6.91, 8.85]	-0.31 [-0.70, 0.11]	0.94 [0.57, 1.33]
Invertivore	8.15 [7.09, 9.18]	8.07 [7.26, 8.89]	-0.48 [-0.81, -0.23]	0.95, 0.78, 1.12]
Nectarivore	7.54 [6.16, 8.74]	7.56 [6.30, 8.64]	-0.13 [-0.53, 0.50]	0.88 [0.36, 1.27]
Omnivore	7.72 [6.64, 8.73]	7.86 [7.05, 8.69]	-0.37 [-0.62, -0.12]	0.93 [0.73, 1.12]
Scavenger	7.77 [6.41, 9.00]	7.95 [6.86, 9.05]	-0.24 [-0.64, 0.41]	0.87 [0.34, 1.22]
Vertivore	8.04 [6.90, 9.14]	8.01 [7.09, 8.96]	-0.32 [-0.68, 0.10]	0.96 [0.62, 1.38]
Body mass	-0.28 [-0.50, -0.07]			
Hand-wing index	0.92 [0.77, 1.05]			
<b>Climatic distance</b>				
(B)	Univariate regression	Multivariate regression		
		Effect	Intercept	Body mass
<b>Trophic guild</b>				
Aquatic predator	1.72 [1.47, 1.96]	1.66 [1.40, 1.90]	-0.03 [-0.11, 0.04]	0.14 [0.05, 0.23]
Frugivore	1.74 [1.48, 2.00]	1.76 [1.48, 2.05]	-0.01 [-0.12, 0.16]	0.01 [-0.18, 0.19]
Granivore	1.72 [1.46, 1.95]	1.71 [1.47, 1.94]	-0.02 [-0.11, 0.06]	-0.01 [-0.12, 0.09]
Herbivore aquatic	1.70 [1.43, 1.95]	1.64 [1.30, 1.93]	-0.01 [-0.10, 0.11]	0.11 [-0.11, 0.34]
Herbivore terrestrial	1.73 [1.47, 1.97]	1.64 [1.30, 1.93]	-0.02 [-0.15, 0.08]	0.21 [-0.04, 0.53]
Invertivore	1.73 [1.48, 1.96]	1.74 [1.50, 1.97]	-0.02 [-0.09, 0.05]	0.05 [-0.01, 0.10]
Nectarivore	1.66 [1.32, 1.94]	1.54 [1.41, 1.89]	0.01 [-0.09, 0.19]	0.32 [0.09, 0.55]
Omnivore	1.72 [1.47, 1.95]	1.73 [1.50, 1.97]	-0.04 [-0.11, 0.02]	0.03 [-0.03, 0.09]
Scavenger	1.66 [1.26, 1.94]	1.58 [1.18, 1.92]	-0.03 [-0.21, 0.10]	0.32 [0.05, 0.63]
Vertivore	1.77 [1.50, 2.03]	1.69 [1.42, 1.97]	-0.01 [-0.12, 0.10]	0.29 [0.08, 0.52]
Body mass	-0.02 [-0.07, 0.03]			
Hand-wing index	0.06 [0.02, 0.10]			



**Figure S2.1: Trade-off between movements in geographic and climatic space after controlling for phylogenetic relatedness of species.** Relationship between climatic and geographic distance travelled by each species. Each point is the sum of distances between the centroids of consecutive weeks in climatic and geographic space weighted by abundance. The line is the mean effects while the envelop is the 95% credible interval built using the posterior predictive distribution of the piecewise model. We highlighted an example of resident ‘niche-switcher’ (Rufous-winged Sparrow), resident ‘niche-tracker’ (Great Kiskadee), migrant ‘niche-switcher’ (Buff-breasted Sandpiper), and a migrant ‘niche-tracker’ (Eastern Wood-Pewee). Turquoise points are the highlighted ‘niche-switchers’, and golden points the highlighted ‘niche-trackers’.



**Figure S2.2: Relationship between migratory behaviour and species traits after for controlling for species phylogenetic relatedness.** Mean effect of body mass (left panels), hand-wing index (centre panels), and trophic guilds (right panels) on geographic distance (upper panels) and climatic distance (lower panels) travel by birds. Each points is a species, and the lines and boxplot are the effects determined by the posterior predictive distribution. For all traits, the figure show the direct effect of the trait (after controlling for other traits).

### Brief interpretation:

Overall, controlling for phylogenetic relatedness did not qualitatively change the direction or size of geographic distance on climatic distance (Fig. S2.1), or the effect of body mass, hand-wing index on geographic and climatic distance travelled by species (Fig. S2.2). One of the main difference is that, after controlling for phylogenetic relatedness, larger species tended to travel less distance in geographic space (without controlling for other traits), whereas we found the opposite relationship when we did not control for phylogeny. Not surprisingly, between-guilds differences are weaker since some of it is explained by phylogeny.

### References:

- Bürkner, P.-C. (2018). Advanced Bayesian Multilevel Modeling with the R Package brms. *The R Journal*, 10(1), 395. <https://doi.org/10.32614/RJ-2018-017>
- Jetz, W., Thomas, G. H., Joy, J. B., Hartmann, K., & Mooers, A. O. (2012). The global diversity of birds in space and time. *Nature*, 491(7424), Article 7424. <https://doi.org/10.1038/nature11631>

Neal, R. M. (2011). MCMC Using Hamiltonian Dynamics. In Handbook of Markov Chain Monte Carlo (pp. 113–162). CRC Press. <https://doi.org/10.1201/b10905-7>

## Appendix S4: Model convergence

### Description:

We evaluated convergence and adequacy of the four predictive models using rank plots, posterior predictive checks, and the rank-normalized potential scale reduction factor on split chains. To keep the appendices within a reasonable length, we only show here the rank-normalized potential scale reduction on split chains (Rhat) which diagnosed convergence, however all models were inspected using rank plots and posterior predictive checks.

Rank trace plots are similar to trace plots and are used to assess the mixing of the chains (Gelman *et al.* 2013). With trace rank plots, we check if the MCMC sampler explores the parameter space adequately. For each iterations (x axis), the parameter varies (y axis), and the chains are ranked from 1 (lowest value of the parameter) to 4 (highest value of the parameter). In the ideal case, we want that the chains to change rank more or less randomly (no chain staying at a given rank more than others). Conversely, if a given chain stays in the same rank for many iterations in a row, or is more often at the same rank than other chains, it is sign of a problem with mixing.

Posterior predictive checks compare the distribution of the observed data to the distribution of the posterior predictive distribution (Gelman *et al.* 2013). A discrepancy in distributions can be caused by a model missfit. A model that fits correctly should replicate the observed distribution of interactions/non interactions (counts on the y axis, and 0 and 1 on the x axis) of each “group” of the model (in our models, we have varying parameters for the order of the predator, so an histogram for each predator order). A problematic model would fail to replicate the distributions.

Finally, the rank-normalized potential scale reduction factor on split chains (Rhat) is similar to the potential scale reduction factor, and is a summary statistics evaluating model convergence (Vehtari *et al.* 2021). Rhat close to 1 is indicate a correctly converged model, and a Rhat larger than 1 ( $Rhat > 1.1$ ) would indicate that the model did not converge correctly.

Here is the list of all models, the associated brms formula and Rhat:

## **Model 1: Picewise model of climate distance as a function of geographic distance**

$$\log(\text{clim\_centroid.dist}) \sim b0 + b1 * (\log1p(\text{geo.dist}) - \text{omega}) * \text{step}(\text{omega} - \log1p(\text{geo.dist})) + b2 * (\log1p(\text{geo.dist}) - \text{omega}) * \text{step}(\log1p(\text{geo.dist}) - \text{omega})$$

Variables and coefficient:

- clim\_centroid.dist is the climatic distance between the breeding and wintering climatic centroid in standard deviation unit
- geo.dist is the grand circle distance between the breeding and winter geographic centroid in km.
- omega is the geographic distance of the breaking point
- b0 is the common intercept
- b1 is the slope before the breaking point
- b2 is the slope after the breaking point

Convergence:

Coefficient	Rhat
B0	1.001
B1	1.000
B2	1.001
omega	1.000

## **Model 2: Linear model of geographic distance as a function of body mass, hand-wing index, and trophic guild**

$$\log1p(\text{geo.dist}) \sim \log\text{BM\_sc} + \log\text{HWI\_sc} + (1 + \log\text{BM\_sc} + \log\text{HWI\_sc}) \cdot \text{Trophic.Niche}$$

Variables:

- geo.dist is the grand circle distance between the breeding and winter geographic centroid in km.
- logHWI\_sc: Scaled and log-transformed hand-wing index
- logBM\_sc: Scaled and log-transformed body mass
- Trophic.Niche: Trophic guild

Convergence:

Coefficient	Rhat
Intercept	1.003
Slope_Body mass	1.004
Slope_hand wing index	1.001
Intercept_Aquatic predator	1.002
Intercept_Frugivore	1.003
Intercept_Granivore	1.001
Intercept_Herbivore aquatic	1.001
Intercept_Herbivore terrestrial	1.001
Intercept_Invertivore	1.002
Intercept_Nectarivore	1.003
Intercept_Omnivore	1.002
Intercept_Scavenger	1.001
Intercept_Vertivore	1.001
Slope_Body mass:Aquatic predator	1.001
Slope_Body mass:Frugivore	1.003
Slope_Body mass:Granivore	1.001
Slope_Body mass:Herbivore aquatic	1.000
Slope_Body mass:Herbivore terrestrial	1.000
Slope_Body mass:Invertivore	1.006
Slope_Body mass:Nectarivore	1.001
Slope_Body mass:Omnivore	1.003
Slope_Body mass:Scavenger	1.000
Slope_Body mass:Vertivore	1.000
Slope_Hand wing index:Aquatic predator	1.001
Slope_Hand wing index:Frugivore	1.000
Slope_Hand wing index:Granivore	1.000
Slope_Hand wing index:Herbivore aquatic	1.000
Slope_Hand wing index:Herbivore terrestrial	1.001
Slope_Hand wing index:Invertivore	1.001
Slope_Hand wing index:Nectarivore	1.000
Slope_Hand wing index:Omnivore	1.000
Slope_Hand wing index:Scavenger	1.000
Slope_Hand wing index:Vertivore	1.000

### Model 3: Linear model of geographic distance as a function of body mass

$$\log1p(\text{geo.dist}) \sim \log\text{BM\_sc}$$

Variables:

- geo.dist is the grand circle distance between the breeding and winter geographic centroid in km.
- logBM\_sc: Scaled and log-transformed body mass

Convergence:

Coefficient	Rhat
Intercept	1.001
Slope_Body mass	1.000

#### Model 4: Linear model of geographic distance as a function of hand-wing index,

$$\log1p(\text{geo.dist}) \sim \log\text{HWI\_sc}$$

Variables:

- geo.dist is the grand circle distance between the breeding and winter geographic centroid in km.
- logHWI\_sc: Scaled and log-transformed hand-wing index

Convergence:

Coefficient	Rhat
Intercept	1.000
Slope_Hand wing index	1.001

#### Model 5: Linear model of geographic distance as a function of trophic guild

$$\log1p(\text{geo.dist}) \sim (1|\text{Trophic.Niche})$$

Variables:

- geo.dist is the grand circle distance between the breeding and winter geographic centroid in km.
- Trophic.Niche: Trophic guild

Convergence:

Coefficient	Rhat
Intercept	1.001
Intercept_Aquatic predator	1.002
Intercept_Frugivore	1.000
Intercept_Granivore	1.002
Intercept_Herbivore aquatic	1.003
Intercept_Herbivore terrestrial	1.002
Intercept_Invertivore	1.006
Intercept_Nectarivore	1.001
Intercept_Omnivore	1.005
Intercept_Scavenger	1.000
Intercept_Vertivore	1.002

## Model 6: Linear model of climatic distance as a function of body mass, hand-wing index, and trophic guild

$$\log(\text{clim\_centroid.dist}) \sim \log\text{BM}_\text{sc} + \log\text{HWI}_\text{sc} + (1+\log\text{BM}_\text{sc}+\log\text{HWI}_\text{sc}| \text{Trophic.Niche})$$

Variables:

- clim\_centroid.dist is the climatic distance between the breeding and wintering climatic centroid in standard deviation unit
- logHWI\_sc: Scaled and log-transformed hand-wing index
- logBM\_sc: Scaled and log-transformed body mass
- Trophic.Niche: Trophic guild

Convergence:

Coefficient	Rhat
Intercept	1.005
Slope_Body mass	1.005
Slope_hand wing index	1.002
Intercept_Aquatic predator	1.004
Intercept_Frugivore	1.004
Intercept_Granivore	1.004
Intercept_Herbivore aquatic	1.001
Intercept_Herbivore terrestrial	1.002
Intercept_Invertivore	1.008
Intercept_Nectarivore	1.002
Intercept_Omnivore	1.004
Intercept_Scavenger	1.004
Intercept_Vertivore	1.003
Slope_Body mass:Aquatic predator	1.002
Slope_Body mass:Frugivore	1.005
Slope_Body mass:Granivore	1.006
Slope_Body mass:Herbivore aquatic	1.001
Slope_Body mass:Herbivore terrestrial	1.001
Slope_Body mass:Invertivore	1.004
Slope_Body mass:Nectarivore	1.006
Slope_Body mass:Omnivore	1.001
Slope_Body mass:Scavenger	1.008
Slope_Body mass:Vertivore	1.001
Slope_Hand wing index:Aquatic predator	1.000
Slope_Hand wing index:Frugivore	1.000
Slope_Hand wing index:Granivore	1.002
Slope_Hand wing index:Herbivore aquatic	1.001
Slope_Hand wing index:Herbivore terrestrial	1.002
Slope_Hand wing index:Invertivore	1.002
Slope_Hand wing index:Nectarivore	1.002
Slope_Hand wing index:Omnivore	1.002
Slope_Hand wing index:Scavenger	1.006

Slope_Hand wing index:Vertivore	1.001
---------------------------------	-------

### Model 7: Linear model of climatic distance as a function of body mass

$$\log(\text{clim\_centroid.dist}) \sim \log\text{BM\_sc}$$

Variables:

- clim\_centroid.dist is the climatic distance between the breeding and wintering climatic centroid in standard deviation unit
- logBM\_sc: Scaled and log-transformed body mass

Convergence:

Coefficient	Rhat
Intercept	1.000
Slope_Body mass	1.000

### Model 8: Linear model of climatic distance as a function of hand-wing index

$$\log(\text{clim\_centroid.dist}) \sim \log\text{HWI\_sc}$$

Variables:

- clim\_centroid.dist is the climatic distance between the breeding and wintering climatic centroid in standard deviation unit
- logHWI\_sc: Scaled and log-transformed hand-wing index

Convergence:

Coefficient	Rhat
Intercept	1.001
Slope_Hand wing index	1.000

### Model 9: Linear model of climatic distance as a function of trophic guild

$$\log(\text{clim\_centroid.dist}) \sim (1|\text{Trophic.Niche})$$

Variables:

- clim\_centroid.dist is the climatic distance between the breeding and wintering climatic centroid in standard deviation unit

- Trophic.Niche: Trophic guild

Convergence:

Coefficient	Rhat
Intercept	1.012
Intercept_Aquatic predator	1.019
Intercept_Frugivore	1.005
Intercept_Granivore	1.008
Intercept_Herbivore aquatic	1.004
Intercept_Herbivore terrestrial	1.013
Intercept_Invertivore	1.012
Intercept_Nectarivore	1.011
Intercept_Omnivore	1.010
Intercept_Scavenger	1.011
Intercept_Vertivore	1.010

### Brief interpretation:

None of the rank-normalized potential scale reduction factor on split chains are larger than 1.02 pointing out to a good convergence of every models. None of the rank plots (not shown here) show a particular systematic pattern in the ranks of the chains suggesting a good mixing of chains for every models. Finally, the distribution of the prediction seems to fit fairly well the distribution of the observed data (posterior predictive checks, not shown here), which suggest that the models are adequate to explain the observed data.

### References:

Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A. & Rubin, D.B. (2013). *Bayesian Data Analysis*. 0 edn. Chapman and Hall/CRC.

Vehtari, A., Gelman, A., Simpson, D., Carpenter, B. & Bürkner, P.-C. (2021). Rank-Normalization, Folding, and Localization: An Improved  $\hat{R}$  for Assessing Convergence of MCMC (with Discussion). *Bayesian Analysis*, 16, 667–718.

## **D. Supplementary material for chapter 4**

**Seasonal level of human impact and protection drive the decline of the North**

**American avifauna**

***Dominique Caron, Pierre Legagneux, David Beauchesne, Phillippe Archambault, Laura J. Pollock***

**Table of Content:**

**Appendix S1 – Multicollinearity sensitivity analysis:** Analysis to detect whether correlation between independent variables might influence the findings

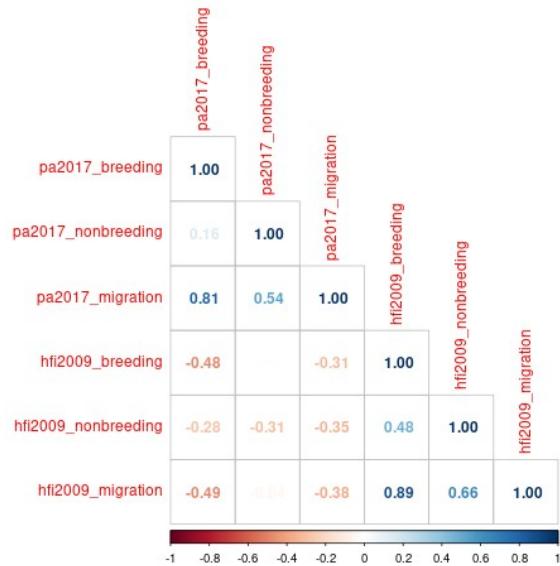
## **Appendix S1 – Multicollinearity sensitivity analysis**

**Description:** In this appendix, we explore the potential influence of the correlations between independent variables for our analysis on the relationship between bird trends and seasonal levels of protection and human impact. Since the distributions of a given species at different seasons is not independent (i.e., birds migrate from its breeding to its nonbreeding range), we might expect strong correlation between the level of protection and human impact at different seasons. Also, the level of human impact and protection is not independent: protected areas should also have lower human impact. Correlation between variables might lead to inflated standard errors of model coefficients (does not seem to be a problem), make the interpretation of the coefficient difficult (address in the discussion), and instability of the coefficient. Here, we will first report the correlation between independent variables of our three models (migratory land birds, migratory water birds, and resident species). Second, we will test the stability of model coefficients. To do this, we will calibrate each model 10 times using 50% of the species (species are randomly drawn each time), and compare model coefficients across the 10 iterations.

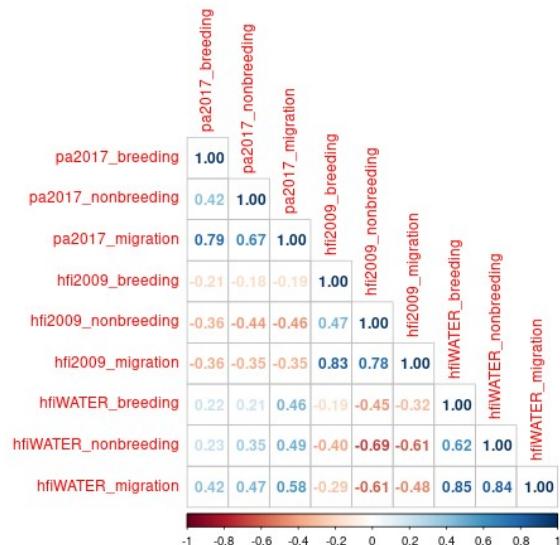
## Results:

*Correlation among independent variables:*

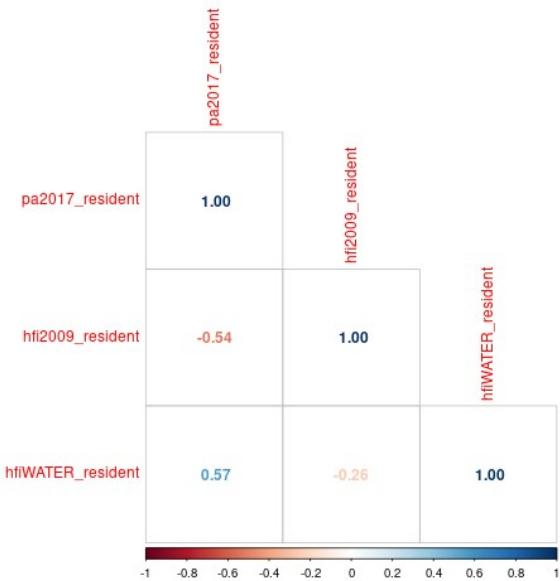
Model 1: Migratory land birds



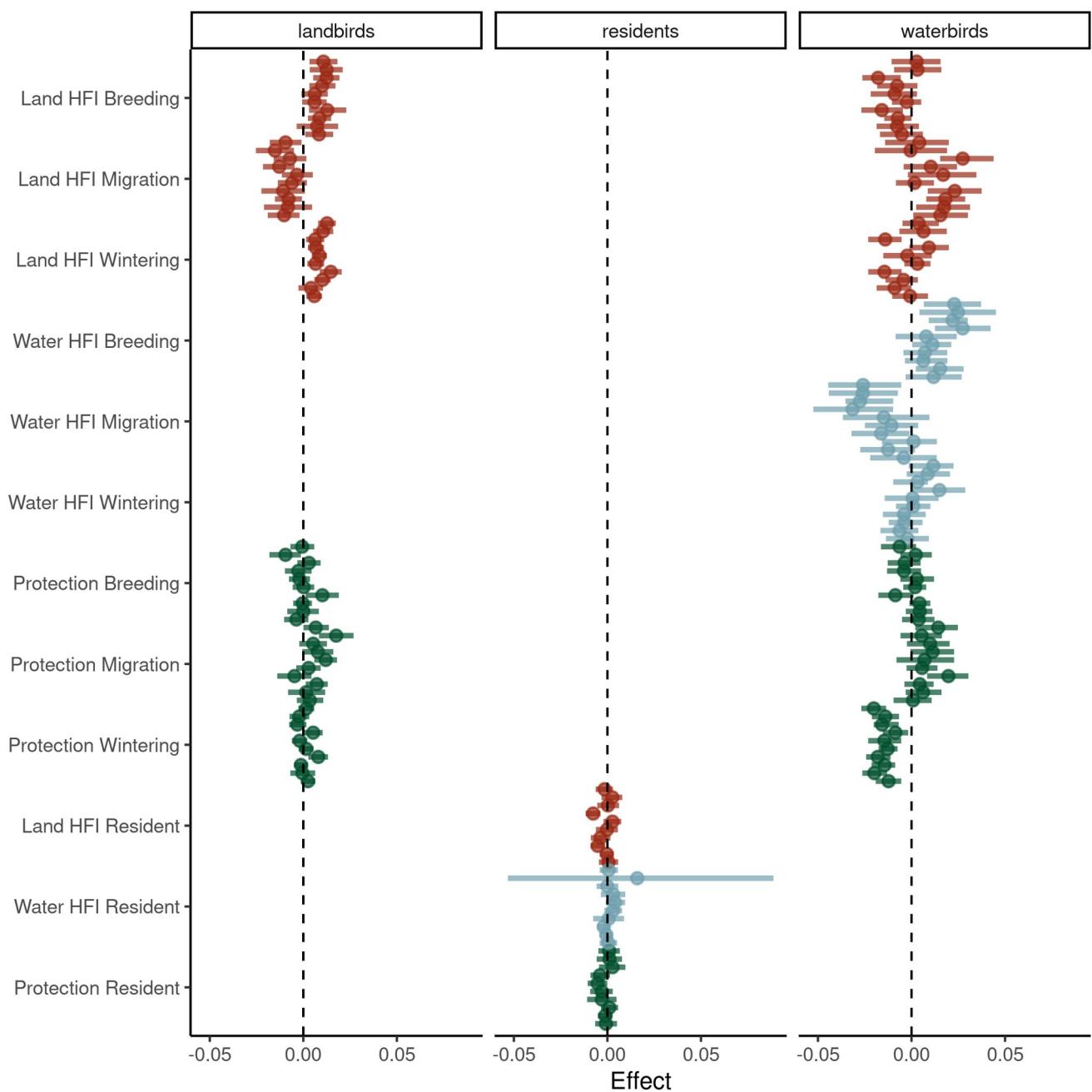
Model 2: Migratory water birds



### Model 3: Resident birds



*Stability of model coefficients:*



**Brief discussion:**

As expected, for a given season, birds on protected areas tend to also be on less impacted land (correlation between HFI and protected area), but the correlation between these variable is not

extremely high ( $r$  less than 0.6). However, the correlation between levels of protection or human impact between different seasons is often high, especially when comparing the breeding and migration season ( $r$  between 0.8 and 0.9). When comparing the model coefficients fitted on 10 different samples of the data, we find that the effect remains qualitatively constant. We found one instance where the standard error of the water hfi variable for the resident bird model was inflated, which is not too surprising given we only have 102, mostly landbird, resident species. Overall, our findings do not seem to be greatly affected by the correlation between independent variables as the main effects found (migration land HFI, breeding land HFI, non breeding land HFI, and non breeding protection on land migratory bird trends; and breeding water HFI, migration water HFI, migration protection, and nonbreeding protection on water migratory bird trends) are stable.